

STIC-Biotech/ChemLib

123121

From: Chan, Christina
Sent: Thursday, May 27, 2004 12:18 PM
To: Schnizer, Holly; STIC-Biotech/ChemLib
Subject: RE: Request for RUSH sequence search in Appl. no. 09/902772

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644 & 1642
(571)-272-0841
Remsen, 3E89

CRFE

-----Original Message-----

From: Schnizer, Holly
Sent: Thursday, May 27, 2004 12:14 PM
To: Chan, Christina
Subject: Request for RUSH sequence search in Appl. no. 09/902772

I would like to request a RUSH sequence search for the sequences below in the above appl. which is an after-final.

Please search all databases for

SEQ ID NO:1
reverse translation of SEQ ID NO:2.

Thank you.

Holly Schnizer
AU 1653
Office: REM 3C79
Mailbox: REM 3C70
(571)272-0958

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 5/28
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 27, 2004, 16:28:06 ; Search time 5614 Seconds
(without alignments)
10787.301 Million cell updates/sec
Title: US-09-902-772-1
Perfect score: 1447
Sequence: 1 gaattccgcgaacgaataat.....gaaagaagcgccaagaaaa 1447
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rtd.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sv.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1447	100.0	1447	6	E31253	E31253 Protein hav
2	1409	97.4	1413	5	AY065661	AY065661 Gallus ga
3	1356	93.7	1528	6	E31254	E31254 Protein hav
4	1344	92.9	1516	5	GGERG	X77159 G.gallus ER
5	994.4	68.7	1509	6	AX657228	AX657228 Sequence
6	953.4	65.9	2016	9	BC040168	BC040168 Homo sapi
7	936	64.7	1808	10	AB031088	AB031088 Rattus no
8	921	63.6	2166	9	HUMERG2	M17254 Human erg2
9	889.6	61.5	2012	5	XLAJ4126	AJ224126 Xenopus 1
10	881.6	60.9	2131	10	AB073078	AB073078 Mus muscu
11	867.2	59.9	2209	10	AB073080	AB073080 Mus muscu
12	833.6	57.6	1291	6	AX657230	AX657230 Sequence
13	828	57.2	2133	10	AB073079	AB073079 Mus muscu
14	801.8	55.4	1798	5	XLAJ4125	AJ224125 Xenopus 1
15	760.2	52.5	3126	9	HUMERG11	M21535 Human erg p
16	742.6	51.3	1372	6	AX657234	AX657234 Sequence
17	698.8	48.3	1219	6	AX657232	AX657232 Sequence
18	571.2	39.5	1436	5	DRE249590	AJ249590 Danio rer
19	570.2	39.4	1359	6	AX695580	AX695580 Sequence
20	570.2	39.4	1729	6	AX695579	AX695579 Sequence
21	570.2	39.4	1729	10	MMFLI1	X59421 Mouse FLI-1
22	563.8	39.0	2916	6	A36461	A36461 Sequence 2
23	563.8	39.0	2938	6	AR080101	AR080101 Sequence
24	562.2	38.9	2388	9	HSUMFLI	X67001 H.sapiens H
25	562.2	38.9	1359	6	AX695583	AX695583 Sequence
26	562.2	38.9	1359	9	AY029368	AY029368 Homo sapi
27	562.2	38.9	2394	9	BC010115	BC010115 Homo sapi
28	562.2	38.9	2957	6	AX587825	AX587825 Sequence
29	562.2	38.9	2957	6	AX695582	AX695582 Sequence
30	562.2	38.9	2957	9	HUMERGFLI	M98833 Homo sapien
31	562.2	38.9	2959	9	BC001670	BC001670 Homo sapi
32	561.4	38.8	6849	12	AY029367	AY029367 Synthetic
33	560.6	38.7	1673	9	S45205	S45205 FLI-1-Frien
34	549.8	38.0	2919	5	BC056588	BC056588 Danio rer
35	549.4	38.0	4403	5	XLFLIG	X66979 X.laevis mr
36	544.6	37.6	2954	6	BD094081	BD094081 Shear str
37	518.6	35.8	1245	5	AF177538	AF177538 Danio rer
38	514	35.5	3490	5	CCRNAPLI	Y14773 Coturnix co
39	502.2	34.7	3545	5	CCFLIONCO	Y14774 Coturnix co
40	493.2	34.1	762	9	S72621	S72621 EWS...erg {
41	478.2	33.0	1932	9	HUMFLI1A	M93255 Human FLI-1
42	477	33.0	816	10	S66169	S66169 Erg-3-immun
43	388.6	26.9	127477	9	BS000128	BS000128 Pan trogl
44	388.6	26.9	148367	9	BS000159	BS000159 Pan trogl
45	387	26.7	14518	9	AP001423	AP001423 Homo sapi

ALIGNMENTS

RESULT 1
E31253
LOCUS
DEFINITION Protein having cell calcifying inhibitory activity and gene
encoding the same.
1447 bp DNA linear PAT 18-JUN-2001
ACCESSION E31253
VERSION E31253.1 GI:13025685
KEYWORDS JP 1999075871-A/1.
SOURCE unidentified
ORGANISM unclassified
REFERENCE 1 (bases 1 to 1447)
AUTHORS Hiroyasu, I., Yoshinobu, H., Marijio, P., Joel, R. and Helena, E.
TITLE Protein having cell calcifying inhibitory activity and gene
encoding the same

Pred. No. is the number of results predicted by chance to have a

Yeh, H., Abrams, W.R., Rosenbloom, J. and Pacifici, M.
Transcription factor ERG variants and functional diversification of
Chondrocytes during limb long bone development
J. Cell Biol. 150 (1), 27-40 (2000)
20351415
JOURNAL MEDLINE
PUBMED 10893254
REFERENCE 2 (bases 1 to 1413)
AUTHORS Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K., Koyama, E.,
Yeh, H., Rosenbloom, J. and Pacifici, M.
TITLE The role of ERG (ets related gene) in cartilage development
Osteoarthritis. Cartil. 9 Suppl A, S41-S47 (2001)
21535378
JOURNAL MEDLINE
PUBMED 11680687
REFERENCE 3 (bases 1 to 1413)
AUTHORS Iwamoto, M., Higuchi, Y., Yeh, H. and Pacifici, M.
TITLE Direct Submission
SUBMITTED (05-DEC-2001) Oral Anatomy and Developmental Biology,
Osaka University Faculty of Dentistry, 1-8 Yamadaoka, Suita, Osaka
565-0871, Japan
FEATURES
source
1. 1413
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
37. 11392
/note="alternatively spliced; transcription factor"
/codon_start=1
/product="erg isoform C-1-1"
/protein_id="AA140889.1"
/db_xref="GI:17887441"
/translation="MSTIKIALSVSEDSLPEYAGSPHLAKTENTASSSEYQGT
SKMSPRVQDNLSPQPARVITKMEFNQVNGSRNSPDDCSVAKGKMVSSSDNVGM
NYGSMYMEKHIPPNTMTNRRVVPADLTWSTHVRQLEWAVKEYGLPVDVILF
QNTDGKELKMTKDDFORLTFSYNADILLSHLYLRERGFIFPNTSVYPEATORIT
TRPLPYEQARRSAWTSHTSPTQSKATQPSSTVPTKEDORPOLDPQYILGPTSSRLA
NPGSGOIQLOFLELLLSNNSNCITWETNGEFKMTDPEVARWGRKSKPNNVY
DKLSRALRYVDKNNIKVGRKAVKDFHGIQAOLQHPHPRESSVKYPSDLPYMS
YHAPQKQNFVAFHPALPVTSSSFPAAPNYWNSPTGGIYFNTRLPAAHMFSLGT
Y"
ORIGIN
Query Match 97.4%; Score 1409; DB 5; Length 1413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
27 TAGCAATTATAGCGATCAATAATCTTGTATGACATTAATGCGACGACCTAATAAGGAAGCA 86
Db 1 TAGCAATTATAGCGATCAATAATCTTGTATGACATTAATGCGACGACCTAATAAGGAAGCA 60
87 TTATCAGTGTGAGTGAAGCAGCTCTTGTTCAGTGTGCTTACGATCGCCACCTT 146
Db 61 TTATCAGTGTGAGTGAAGCAGCTCTTGTTCAGTGTGCTTACGATCGCCACCTT 120
147 GCAAGACAGAAATGACAGCCTCTCTTCCAGTGAATATATGGGCAACATCAAGATGAGC 206
Db 121 GCAAGACAGAAATGACAGCCTCTCTTCCAGTGAATATATGGGCAACATCAAGATGAGC 180
207 CCGCGCTTCCAGCAGGAGTGTATACAGCCCGCGGCGAGATTACCATTAAGATG 266
Db 181 CCGCGCTTCCAGCAGGAGTGTATACAGCCCGCGGCGAGATTACCATTAAGATG 240
267 GAGTGTAAACCAACACAGGTTAAATGGGTCAAGGAATTCACCTGATGACTGCGAGCTGGCA 326
Db 241 GAGTGTAAACCAACACAGGTTAAATGGGTCAAGGAATTCACCTGATGACTGCGAGCTGGCA 300
327 AAAGGAGGAAATGGTTAGCAGTTCAGACAAATGTTGGATGACTGAGAGCTACATG 386
Db 301 AAAGGAGGAAATGGTTAGCAGTTCAGACAAATGTTGGATGACTGAGAGCTACATG 360
387 GAAGAGAGCATATTTCCGCTTCCAAATATGACCAATGAACGAAGATTATTGTGCCA 446
Db 361 GAAGAGAGCATATTTCCGCTTCCAAATATGACCAATGAACGAAGATTATTGTGCCA 420

447 CGAGATCCTACGTTTATGAGCAGACACCATGTACGGCAGTGGCTGGAGTGGGCACTGAAG 506
Db 421 CGAGATCCTACGTTTATGAGCAGACACCATGTACGGCAGTGGCTGGAGTGGGCACTGAAG 480
507 GAGTATGCTTCCAGACGTCGACATCTTGTGTTCAGAAACATTTGATGGGAAAGAGTTG 566
Db 481 GAGTATGCTTCCAGACGTCGACATCTTGTGTTCAGAAACATTTGATGGGAAAGAGTTG 540
567 TGTAAATGACCAAGATGACTTCCAGAGACTCAGCGGAGCTATAACGCGAGATATCCTC 626
Db 541 TGTAAATGACCAAGATGACTTCCAGAGACTCAGCGGAGCTATAACGCGAGATATCCTC 600
627 CTGTCCACACTACACTACCTCAGAGAGAGAGGACCACTTTTATTTTCCAAATACATCA 686
Db 601 CTGTCCACACTACACTACCTCAGAGAGAGAGGACCACTTTTATTTTCCAAATACATCA 660
687 GTTTACCCAGAGACCAACGCAAGATACACAGGCGAGATTTACCTTATGAGCAAGCG 746
Db 661 GTTTACCCAGAGACCAACGCAAGATACACAGGCGAGATTTACCTTATGAGCAAGCG 720
747 AGGAGATCAGCGTGGACGAGTCACAGCCATCCCACTCAGTCAAAAGCTACCCCAACATCA 806
Db 721 AGGAGATCAGCGTGGACGAGTCACAGCCATCCCACTCAGTCAAAAGCTACCCCAACATCA 780
807 TCTTCAACAGTCCCAAAACAGAGACGAGTCTCTCAGTTAGATCCTTATCAGATTCCT 866
Db 781 TCTTCAACAGTCCCAAAACAGAGACGAGTCTCTCAGTTAGATCCTTATCAGATTCCT 840
867 GGACCGACAGCAGCGCTCTTGCATTCAGAGGAGTGGGCGAGATACAGTATGGCAGTTC 926
Db 841 GGACCGACAGCAGCGCTCTTGCATTCAGAGGAGTGGGCGAGATACAGTATGGCAGTTC 900
927 CTATCGAGCTTCTGTGGAAGCTCCAACTCAACCTCAACCTGATCCTTGGAGGGGCAAAAT 986
Db 901 CTATCGAGCTTCTGTGGAAGCTCCAACTCAACCTGATCCTTGGAGGGGCAAAAT 960
987 GGGGAGTTCAAGATCAGACAGCCTCATGAAGTGGCTCGCGTGGGGAGAGAGGAAAGC 1046
Db 961 GGGGAGTTCAAGATCAGACAGCCTCATGAAGTGGCTCGCGTGGGGAGAGAGGAAAGC 1020
1047 AAACCTAACATGAATATGACAACTCAGCGCTGCACTTGGCTACTACTATGACAAAT 1106
Db 1021 AAACCTAACATGAATATGACAACTCAGCGCTGCACTTGGCTACTACTATGACAAAT 1080
1107 ATTATGACTAAAGTTTCATGTTAAAGCTATGCTCAAAATTTGATTTCCACGGAATCGCT 1166
Db 1081 ATTATGACTAAAGTTTCATGTTAAAGCTATGCTCAAAATTTGATTTCCACGGAATCGCT 1140
1167 CAGGCGCTCCAGCTCAGCTCCAGATCATCCATGTACAAATACCCATCAGACCTCCCT 1226
Db 1141 CAGGCGCTCCAGCTCAGCTCCAGATCATCCATGTACAAATACCCATCAGACCTCCCT 1200
1227 TACATGAGTTCCTACCATGCACACCCCGAGAGATGAATTTGTAGTCCCATCCCT 1286
Db 1201 TACATGAGTTCCTACCATGCACACCCCGAGAGATGAATTTGTAGTCCCATCCCT 1260
1287 GTTTGGCCCGTAACTCATCCAGCTTTTGTGCTCCCTTAATCCATCTGGAATCAACA 1346
Db 1261 GTTTGGCCCGTAACTCATCCAGCTTTTGTGCTCCCTTAATCCATCTGGAATCAACA 1320
1347 ACTGAGGATCATCCCAATACAGGCTCCAGTGTCTCATATGCTTCCCTTCCCTTGGC 1406
Db 1321 ACTGAGGATCATCCCAATACAGGCTCCAGTGTCTCATATGCTTCCCTTCCCTTGGC 1380
1407 ACTTACTTAAAGTGGGAAAGAAAGAAA 1435
Db 1381 ACTTACTTAAAGTGGGAAAGAAAGAAA 1409
RESULT 3
E31254
LOCUS
DEFINITION Protein having cell calcifying inhibitory activity and gene

1528 bp DNA linear PAR 19-JUN-2001
E31254
LOCUS
DEFINITION Protein having cell calcifying inhibitory activity and gene


```

encoding the same.
E31254
VERSION E31254.1 GI:13025686
KEYWORDS JP 1999075871-A/2.
SOURCE unidentified
ORGANISM unclassified.
1. (bases 1 to 1528)
REFERENCE Hiroyasu,I., Yoshinobu,H., Marijo,P., Joel,R. and Helena,E.
AUTHORS Protein having cell calcifying inhibitory activity and gene
TITLE encoding the same
JOURNAL Patent: JP 1999075871-A 2 23-MAR-1999;
CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
PENNSYLVANIA
OS Unidentified
PN JP 1999075871-A/2
PD 23-MAR-1999
PF 29-MAY-1998 JP 1998166076
PR 18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PI
HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJO PASHIFIKI, PI JOEL
ROZENBUROOM
PI HELENA E
PC C12M15/09,A61K48/00,C07K14/465,C07K16/18,C12Q1/68//A61K38/00,
C12P21/02,
PC C12P21/08,C12M15/00,A61K37/02
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..1528
FT /organism="Unidentified".
FEATURES
source
1..1528
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 93.7%; Score 1356; DB 6; Length 1528;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1447; Conservative 0; Mismatches 0; Indels 81; Gaps 1;
QY 1 GAATTCGCGAAGCAATTAATTATTAGCAATTAATTAGCGATCAATAATCTTGATCACA 60
DB 1 GAATTCGCGAAGCAATTAATTATTAGCAATTAATTAGCGATCAATAATCTTGATCACA 60
QY 61 TTATGGCAAGCACTATTAAAGGAAGCAATTAATCAAGTGGTGAAGACCAAGTCTTGTG 120
DB 61 TTATGGCAAGCACTATTAAAGGAAGCAATTAATCAAGTGGTGAAGACCAAGTCTTGTG 120
QY 121 AGTGTGCTACGGATCGCCCACTTCGAAAGCAGAAATGACAGCTCTCTTCCAGTG 180
DB 121 AGTGTGCTACGGATCGCCCACTTCGAAAGCAGAAATGACAGCTCTCTTCCAGTG 180
QY 181 AATATGGCAACATCAAGATGAGCGCGGTTCCCGAGCAGACTGGTTATCAAGC 240
DB 181 AATATGGCAACATCAAGATGAGCGCGGTTCCCGAGCAGACTGGTTATCAAGC 240
QY 241 CCCGCGCAGAGTTACATTAAGATGAGTGTAACCAACCAACAGGTTAATGGGTCAAGGA 300
DB 241 CCCGCGCAGAGTTACATTAAGATGAGTGTAACCAACCAACAGGTTAATGGGTCAAGGA 300
QY 301 ATTCACTGATGACTGCAAGCTGGCAAGAGGAAATGTTAGCAGTTCAGACAATG 360
DB 301 ATTCACTGATGACTGCAAGCTGGCAAGAGGAAATGTTAGCAGTTCAGACAATG 360
QY 361 TTGGATGAATATGGAAGTACATGGAAGAGAGCATATTCGCTCCCAATATGACAA 420
DB 361 TTGGATGAATATGGAAGTACATGGAAGAGAGCATATTCGCTCCCAATATGACAA 420
QY 421 CCAATGACGAGAGTATTGTCAGCAGATCTTACGTTATGAGCAGACCATGTAC 480
DB 421 CCAATGACGAGAGTATTGTCAGCAGATCTTACGTTATGAGCAGACCATGTAC 480

```

```

481 GCAGTGGCTGGAGTGGCAGTGAAGAGTATGTTCTTCCAGAGTGGACATCTTGTGT 540
481 GCAGTGGCTGGAGTGGCAGTGAAGAGTATGTTCTTCCAGAGTGGACATCTTGTGT 540
541 TCCAGAACATTGATGGGAAAGAGTTGTATAAATGACCAAAAGATGACTTCCAGAGACTCA 600
541 TCCAGAACATTGATGGGAAAGAGTTGTATAAATGACCAAAAGATGACTTCCAGAGACTCA 600
601 GCGCAGCTATTAACGAGATATCTCTCTGTGACAGCTACATCTACCTCAGAGAGA----- 654
601 GCGCAGCTATTAACGAGATATCTCTCTGTGACAGCTACATCTACCTCAGAGAGACTCTC 660
655 -----
661 TTCCACATTTGACTTCAGATGTTGATAAGGCTTACAAAATCTCTCCCGTTAATGC 720
655 -----GAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAG 699
721 ATGCTAGAAAACACAGGAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAG 780
700 CAAACGAAAGATAACAAAGGCCAGATTTACCTTTATGAGCAAGCGAGAGATCAGCGT 759
781 CAAACGAAAGATAACAAAGGCCAGATTTACCTTTATGAGCAAGCGAGAGATCAGCGT 840
760 GGACGAGTCAAGCCATCCCACTCAGTCAAAAGCTTACCCAACTATCTTCAACAGTGC 819
841 GGACGAGTCAAGCCATCCCACTCAGTCAAAAGCTTACCCAACTATCTTCAACAGTGC 900
820 CCAAAACAGAGACCAAGCTCTCAGTTAGATCCTTATCAGATCTTCCGAGCGACCAACA 879
901 CCAAAACAGAGACCAAGCTCTCAGTTAGATCCTTATCAGATCTTCCGAGCGACCAACA 960
880 GCGCTCTTGCAAAATCCAGGAGTGGGAGATACAGCTATGCGAGTTCCTACTGGAGCTTC 939
961 GCGCTCTTGCAAAATCCAGGAGTGGGAGATACAGCTATGCGAGTTCCTACTGGAGCTTC 1020
940 TGTCCGAGCTCCAACTCCAACTGATCACTGCGGAGGAGCAAAATGGGAGTTCAAGA 999
1021 TGTCCGAGCTCCAACTCCAACTGATCACTGCGGAGGAGCAAAATGGGAGTTCAAGA 1080
1000 TGACAGACCTGATGAAGTGGCTCGGCGTTGGGAGAGAGAGAAAGCAAACTAAACATGA 1059
1081 TGACAGACCTGATGAAGTGGCTCGGCGTTGGGAGAGAGAGAAAGCAAACTAAACATGA 1140
1060 ACTATGACAAACTCAGCGTGCCTTCTGCTACTACTATGACAAATAATTTATGACTAAAG 1119
1141 ACTATGACAAACTCAGCGTGCCTTCTGCTACTACTATGACAAATAATTTATGACTAAAG 1200
1120 TTTCATGGTAAACGCTATGCTACAAATTTGATTTTCCAGGAATCGCTCAGGCGCTCCAGC 1179
1201 TTTCATGGTAAACGCTATGCTACAAATTTGATTTTCCAGGAATCGCTCAGGCGCTCCAGC 1260
1180 CTCACCTCCAGAACTCATGTAATAATCCGATCAGCTCCGCTACATGAGTTCT 1239
1261 CTCACCTCCAGAACTCATGTAATAATCCGATCAGCTCCGCTACATGAGTTCT 1320
1240 ACCATGCAACACCCCGAGAGATGAATTTGTAGCTTCCCATCCCTGCTTTGCCCGTAA 1299
1321 ACCATGCAACACCCCGAGAGATGAATTTGTAGCTTCCCATCCCTGCTTTGCCCGTAA 1380
1300 CTTCTATCAGCTTTTGTGTCCTTATCCATCTGGAATTCACCAACTGGAGGATCT 1359
1381 CTTCTATCAGCTTTTGTGTCCTTATCCATCTGGAATTCACCAACTGGAGGATCT 1440
1360 ACCCAATATCAGGCTGCCAGCTGCTATGCTCTTCCATCTTGGCAGCTACTACTAAG 1419
1441 ACCCAATATCAGGCTGCCAGCTGCTATGCTCTTCCATCTTGGCAGCTACTACTAAG 1500
1420 TGGGGAAGAAAGAGCGCCCAAGAAA 1447
1501 TGGGGAAGAAAGAGCGCCCAAGAAA 1528

```

RESULT 4

GGERG
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

G.gallus ERG mRNA.
X77159
X77159.1 GI:790439
ERG gene.
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

1
Dordain, P., Dewitte, F., Desbians, X., Stehelin, D. and
Duterque-Coquillaud, M.
Mesodermal expression of the chicken erg gene associated with
precartilaginous condensation and cartilage differentiation
Mech. Dev. 50 (1), 17-28 (1995)
95329425
7605748
2 (bases 1 to 1516)
Duterque-Coquillaud, M.
Direct Submission
Submitted (17-JAN-1994) M. Duterque-Coquillaud, CNRS UA 1160,
Oncologie Moléculaire, Institut Pasteur, 1 rue Calmette, 59019
Lille, FRANCE

FEATURES

Location/Qualifiers

1..1516
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/tissue types="spleen"
/clone_lib="lambda gt10"
/dev_stage="adult"
1..1516
/gene="ERG"
63..1499
/gene="ERG"
/codon_start=1
/protein_id="CAA54404.1"
/db_xref="GI:790440"
/db_xref="GOA:Q90837"
/db_xref="SWISS-PROT:Q90837"
/translation="MASTIKALSVVSEDSQSLFECAYGSPHLAKTEMTASSSEYQQT
SKNSPRVQDMLSQSPARVITIKMECPNQNVRNSRNSPDSCVAKGKRMVSSSDNVGM
NYGSMYBEXHIPPNNITDRLVWADPTLWSTHDVROWLEWAVEYGLPVDVILF
ONTIGKELCWTKDDFORLTPSYNADILASHLYLRETLPLHLTSDVDKALQNSPL
MHRNKGGAFTIPNPSVYEATQRTITREDLPYEQARRASANTSHSPQSKATQPSL
STVPEKEDQRPDLPYQILGPTSSRLANPGSGQIQWQLELLELSDSSNSNCITWEGT
NGEFKMTDPEVARRWGERKSKPNNDYKLRALRYDDKNTIMTKYHGRYAKYKDFH
GIAAQLPHPPESMYKPSDLPYNGSYFAHPQKNFVAPHPPELFPVTSSEFFAAPNP
YMSPTGGIYVENTRLPAAHMPSHLGTYY"

ORIGIN

Query Match 92.9%; Score 1344; DB 5; Length 1516;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1435; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

QY 1 GAATTCGCGAAGCAATTAATTATTAGCAATTAATTAGCGATCAATAATCTTGATCACA 60
DB 1 GAATTCGCGAAGCAATTAATTATTAGCAATTAATTAGCGATCAATAATCTTGATCACA 60

QY 61 TTATGGCAAGCACTATTAAAGAGCACTATCAGTGGTGAAGACCACTCTTGTGTTG 120
DB 61 TTATGGCAAGCACTATTAAAGAGCACTATTATCAGTGGTGAAGACCACTCTTGTGTTG 120

QY 121 AGTGTGCTACGATCGCCCACTTCGAAGACAGAAATGACAGCCTCTCTTCAGTG 180
DB 121 AGTGTGCTACGATCGCCCACTTCGAAGACAGAAATGACAGCCTCTCTTCAGTG 180

QY 181 AATATGGCAAAACATCAAAAGATAGCCGCGGTTCGCCAGAGGAGCTGTTATCAAGC 240
DB 181 AATATGGCAAAACATCAAAAGATAGCCGCGGTTCGCCAGAGGAGCTGTTATCAAGC 240

QY 241 CCCCGGCAGAGTTACCATTAAGATGGAGTGTAAACCCAAACCCAGGTTAATGGGTCAAGA 300
DB 241 CCCCGGCAGAGTTACCATTAAGATGGAGTGTAAACCCAAACCCAGGTTAATGGGTCAAGA 300

QY 301 ATTACCTGATGACTGACGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTTCAGACAATG 360
DB 301 ATTACCTGATGACTGACGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTTCAGACAATG 360

QY 361 TTGGGATGAACCTATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420
DB 361 TTGGGATGAACCTATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420

QY 421 CCAATGAACGAAGAGTTATTGTGCCAGCAGATCTCTACGTTATGGAGCAGACCAATGTAC 480
DB 421 CCAATGAACGAAGAGTTATTGTGCCAGCAGATCTCTACGTTATGGAGCAGACCAATGTAC 480

QY 481 GGCAGTGGCTGGAGTGGCAGTGAAGAGTATGGTCTTCAGACGTCGACATCTTGTTGT 540
DB 481 GGCAGTGGCTGGAGTGGCAGTGAAGAGTATGGTCTTCAGACGTCGACATCTTGTTGT 540

QY 541 TCCAGAACATTGATGGGAAAGAGTTGTGTAATAATGAACAAAGATGACTTCCAGAGACTCA 600
DB 541 TCCAGAACATTGATGGGAAAGAGTTGTGTAATAATGAACAAAGATGACTTCCAGAGACTCA 600

QY 601 CGCCGAGCTATAACGACATATCTCTCTGTCACACCTACACTACCTCAGAGAGA----- 654
DB 601 CGCCGAGCTATAACGACATATCTCTCTGTCACACCTACACTACCTCAGAGAGACTCCTC 660

QY 655 ----- 654
DB 655 ----- 654

QY 661 TTCCACATTGTGACTTTCAGATGATGTTGATAAGCGCTTACAAACCTCTCCACGGTTAATGC 720
DB 661 TTCCACATTGTGACTTTCAGATGATGTTGATAAGCGCTTACAAACCTCTCCACGGTTAATGC 720

QY 655 -----GAGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAG 699
DB 655 -----GAGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAG 699

QY 721 ATGCTAGAAACACAGGAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAG 780
DB 721 ATGCTAGAAACACAGGAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAG 780

QY 700 CAAACCAAGAAATAACAACAAAGGCGAGATTACCTTATGAGCAAGCGAGAGATCAGCGT 759
DB 700 CAAACCAAGAAATAACAACAAAGGCGAGATTACCTTATGAGCAAGCGAGAGATCAGCGT 759

QY 781 CAAACCAAGAAATAACAACAAAGGCGAGATTACCTTATGAGCAAGCGAGAGATCAGCGT 840
DB 781 CAAACCAAGAAATAACAACAAAGGCGAGATTACCTTATGAGCAAGCGAGAGATCAGCGT 840

QY 760 GGAAGAGTACAGCCATCCCACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGC 819
DB 760 GGAAGAGTACAGCCATCCCACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGC 819

QY 841 GGACGAGTCACAGCCATCCCACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGC 900
DB 841 GGACGAGTCACAGCCATCCCACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGC 900

QY 820 CCAAAACAGAAAGACAGAGTCTCAGTTAGATCCTTATCAGATTCTTGACCGACACGCA 879
DB 820 CCAAAACAGAAAGACAGAGTCTCAGTTAGATCCTTATCAGATTCTTGACCGACACGCA 879

QY 901 CCAAAACAGAAAGACAGAGTCTCAGTTAGATCCTTATCAGATTCTTGACCGACACGCA 960
DB 901 CCAAAACAGAAAGACAGAGTCTCAGTTAGATCCTTATCAGATTCTTGACCGACACGCA 960

QY 880 GCGCTCTTTGCAAAATCCAGGAGTGGCGAGATACAGCTATGCGAGTTCTTCTGAGGCTTC 939
DB 880 GCGCTCTTTGCAAAATCCAGGAGTGGCGAGATACAGCTATGCGAGTTCTTCTGAGGCTTC 939

QY 940 TGTCCGACAGCTCCAACTCCAACTCCATCCTGGAGGGGCAAAATGGGAGTTCCAGA 999
DB 940 TGTCCGACAGCTCCAACTCCAACTCCATCCTGGAGGGGCAAAATGGGAGTTCCAGA 999

QY 1021 TGTCCGACAGCTCCAACTCCAACTCCATCCTGGAGGGGCAAAATGGGAGTTCCAGA 1080
DB 1021 TGTCCGACAGCTCCAACTCCAACTCCATCCTGGAGGGGCAAAATGGGAGTTCCAGA 1080

QY 1000 TGACAGACCCCTGATGAGTGGCTCGCGCTTGGGAGAGAGGAAAGAAACCTTAACTGA 1059
DB 1000 TGACAGACCCCTGATGAGTGGCTCGCGCTTGGGAGAGAGGAAAGAAACCTTAACTGA 1059

QY 1081 ACTATGACAAACTCAGCGTGCACCTTCCGCTACTACTATGACAAATAATATTATGACTAAG 1119
DB 1081 ACTATGACAAACTCAGCGTGCACCTTCCGCTACTACTATGACAAATAATATTATGACTAAG 1119

QY 1141 ACTATGACAAACTCAGCGTGCACCTTCCGCTACTACTATGACAAATAATATTATGACTAAG 1200
DB 1141 ACTATGACAAACTCAGCGTGCACCTTCCGCTACTACTATGACAAATAATATTATGACTAAG 1200

QY 1120 TTCTGTTAAAGCTTATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGC 1179
DB 1120 TTCTGTTAAAGCTTATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGC 1179

QY 1201 TTCTGTTAAAGCTTATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGC 1260
DB 1201 TTCTGTTAAAGCTTATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGC 1260

QY 1180 CTCACCTCCAGAAATCATCCATGTACAAATACCCATCAGCTCCCTCCCTACATGCTCCT 1239
DB 1180 CTCACCTCCAGAAATCATCCATGTACAAATACCCATCAGCTCCCTCCCTACATGCTCCT 1239

QY 1261 CTCACCTCCAGAAATCATCCATGTACAAATACCCATCAGCTCCCTCCCTACATGCTCCT 1320
DB 1261 CTCACCTCCAGAAATCATCCATGTACAAATACCCATCAGCTCCCTCCCTACATGCTCCT 1320

```
OY 1240 ACCATGACACCCCGAGAGTGAACCTTTGTAGTCTCCCATCCCTGCTTTCCTGCGTAA 1299
Db 1321 ACCATGACACCCCGAGAGTGAACCTTTGTAGTCTCCCATCCCTGCTTTCCTGCGTAA 1380
OY 1300 CCTCATCAGCTTTTGTGCGCCCTAATCATCTAGTGAATTCACCACTGGAGGATCT 1359
Db 1381 CCTCATCAGCTTTTGTGCGCCCTAATCATCTAGTGAATTCACCACTGGAGGATCT 1440
OY 1360 ACCCAATACAGAGCTGCAGCTGCTCATATGCTTCCCATCTGGACCTACTACTAAG 1419
Db 1441 ACCCAATACAGAGCTGCCAGTCTCATATGCTTCCCATCTGGACCTACTACTAAG 1500
OY 1420 TGGGAAAGAAAGAAA 1435
Db 1501 TGGGAAAGAAAGAAA 1516

RESULT 5
AX657228
LOCUS AX657228 1509 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 1 from Patent WO03000724.
ACCESSION AX657228
VERSION AX657228.1 GI:29160006
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Luyten,F., de Bari,C. and Dell'Accio,F.
TITLE Polynucleotide sequences and vectors useful for the prevention or
treatment of bone- or cartilage-related disorders
JOURNAL Patent: WO 03000724-A 1 03-JAN-2003;
Tigenix N.V. (BE)
FEATURES
source Location/Qualifiers
1..1509
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
67..1446
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD79711.1"
/db_xref="GI:29160007"
/db_xref="REMBL:CAD79711"
/translation="MIOTVPDAHLKEALSUVSDQSLPECAVGTPLHAKTMTASS
SSDYQTSKMSRPVQDNLISQPPARTVIMKCNPSQVNGSRNPSDECSVAKGKMG
SPDVTGMNYSIMEEKHPFNMTNERRVIVPADPLMSTDRVROWLEWAVKEYGLP
EATQITRPDLPEYPPRSAGTGHGHTPQSKAOFSESTVPKTEDQRPOLDPYQIL
GPTSSRLANFGSGQIQWLQFLLELLSDSSNSCITWEGTNGEFKMTDPOEVARRGER
KSKPMNVDKLSRALRYVDKNTMKVHGKRYAYKPDFHGIAQALPHPPESLYKYP
SDLPYMGSYHAHPQKNFVAPHPALPVTSSFFAAPNPVWNSPTGGIYPNTRLPISH
MPSHLGITYY"
CDS
Query Match 68.7%; Score 994.4; DB 6; Length 1509;
Best Local Similarity 84.1%; Pred. No. 7.9e-290;
Matches 1134; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

OY 74 TATTAAAGAACATTATCNGTGGTGAAGTGAAGACGACCTCTCTTCCAGTGAATATGGCAAAAC 133
Db 99 TATCAAGAGACCTTATCAGTTGTGAGTGAGGACGACGTCGTGTGTTGAGTGTGCGCTACGG 158
OY 134 ATCGCCCCACCTTGCAAGACAGAAATGACAGCCTCTCTTCCAGTGAATATGGCAAAAC 193
Db 159 AACGCCACCTGGCTAAGACAGAGATGACCGCGTCTCTCTCCAGGACATATGACAGAC 218
OY 194 ATCAAGATAGCCGCGCTTCCAGAGGAGTGTGTTATCAGAGCCCCCGGCGAGGT 253
Db 219 TTCCAAGATAGCCACGCGTCCCTCAGCAGGATTGGCTGTCTCAACCCCCCAGCCAGGT 278
OY 254 TACCATTAAAGATGAGTGTAAACCAACAGGTGTAATGGGTCAAGGAATTCACCTGATGA 313
```

```
Db 279 CACCATCAAAATGGATGTAACCTTAGCCAGGTGAATGGCTCAAGGAACCTCTCCTGATGA 338
OY 314 CTGAGAGCTGGCAAAAGAGGAAAATGGTTAGCAGTTTCAGACATGTTGGGATGAACATA 373
Db 339 ATGCAAGTGTGGCAAAAGCGGGAAGATGGTGGGAGCCCAACACACCGTGTGGATGAACATA 398
OY 374 TGGAAAGCTACATGGAAGAGAAAGCATATTCGCGCTTCCAAATATGACAACTAATGAACGAG 433
Db 399 CGGAGCTACATGGAAGAGAAAGCATATGCAACCCCAACATGACCAAGAGCGGAG 458
OY 434 AGTTATTTGCGCAGAGATCTTACGTTATGAGACAGACACCATGTACGGCAGTGGCTGGA 493
Db 459 AGTTATTTGCGCAGAGATCTTACGTTATGAGACAGACACCATGTACGGCAGTGGCTGGA 518
OY 494 GTGGGAGTGAAGAGATGATGGTCTTCCAGAGCGTGACATCTTGTGTTCCAGAAACATGA 553
Db 519 GTGGGCGGTGAAGAAATATGGCTTCCAGAGCGTCAACATCTTGTATTCAGAAACATGA 578
OY 554 TGGGAAAGAGTTGTGTAATAATGACCAAAAGATGATCTTCAGAGATCTCAGCGGAGTATAA 613
Db 579 TGGGAAAGAGTTGTGTAATAATGACCAAAAGATGATCTTCAGAGAGGCTCACCCAGCTACAA 638
OY 614 CGCAGATATCTCTCTGTCACACCTACACTACTCAGAGAGAGAGGACACTTTTATTTT 673
Db 639 CGCCGACATCTCTCTCTCAGATCTCCACTACTCAGAGAGAGGAGTGGAGCTTTTATTTT 698
OY 674 TCCAAATACATCAGTTTACCAGAGCAACCAAGAAATAACAACAGCCAGATTTACC 733
Db 699 CCAAAATATCTCAGTATATCTGGAAGCTACCAAGAGATTAACAATAGGCCAGATTTACC 758
OY 734 TTATGAGCAAGCGGAGGATCAGCGTGAACGAGTACAGAGCCATCCCA--CTCAGTCAAA 790
Db 759 ATATGAGCCCCCAGAGATCAGCGCTGAGCGGTACGCGCCACCCAGCCCGGAGTCCGAA 818
OY 791 AGCTACCAACCATCATCTTCAACAGTGCCTCAAAACAGAAAGACAGCGCTCTCTCAGTTAGA 850
Db 819 AGCTGCTCAACCATCTCTTCCACAGTGCCTCAAAACAGAAAGACAGCGCTCTCTCAGTTAGA 878
OY 851 TCCTTATCAGATCTTGGACCGACCGAGCGGTCTTCAAAATCCAGAGAGTGGGAGAT 910
Db 879 TCCTTATCAGATCTTGGACCGACCAAGTAGCGCGCTTCAAAATCCAGAGAGTGGGAGAT 938
OY 911 ACAGTATGCGAGTCTCTACTTGGAGCTTCTGTCGACAGCTCCAACTCCAACTCAGCTCAC 970
Db 939 CCAGCTTTGGCAGTTCTCTCTGAGAGTCTCTGTCGACAGCTCCAACTCAGCTCAC 998
OY 971 CTGGAGGAGCAAAATGGGAGTTCAGATGACAGACCTGATGAGTGGCTCGCGCTTG 1030
Db 999 CTGGGAAAGGCAACACGGGAGTTCAGATGACGATCCCGAGAGTGGGCGCGCGCTG 1058
OY 1031 GGGAGAGAGGAAAAGCAACCTTAACATGAACTATGACAAACTCAGCGCTGACATTCGGCTA 1090
Db 1059 GGGAGAGCGGAAGAGCAAAACCAACATGACTACGATAAGCTCAGCGCGCTCCGTTA 1118
OY 1091 CTACTATGACAAAATATATATGACTAAAGTTTCATGTTAAACGCTATGCTACAAATTTGA 1150
Db 1119 CTACTATGACAAAACATCATGACCAAGGTCATGGAAGCGCTACAGCTCAAGTTGGA 1178
OY 1151 TTTCCAGGAACTCGCTCAGGCGCTCCAGCTCACCTCCAGAACTCATCATGTACAAATA 1210
Db 1179 CTTCACGGGATCGCCGAGGCGCTCCAGCCCAACCCCGGAGTCACTCTGTACAAGTA 1238
OY 1211 CCGATCAGACTCTCCCTTACATGAGTTCTTACATGCAACCCCGAGAGATGAATTTGT 1270
Db 1239 CCGCTCAGACTCTCCCTTACATGAGTTCTTACATGCAACCCCGAGAGATGAATTTGT 1298
OY 1271 AGCTCCCATCCCTGCTTTGCCGTAACCTCATCCAGCTTTTGTGCGCCCTTAATCC 1330
Db 1299 GGGCGCCCACTCCCTGCGCTCCCGTCACTCTCCAGTCTTTTGTGCGCCCAACCC 1358
OY 1331 ATACTGGAATTCACCAACTGGAGGATCTAOCCTAATACAGGCTGCGAGCTGTCTATAT 1390
```

```

Db      1359 ATACTGGAATTCACAACTGGGGGTATATACCCACACTAGGCTCCACCAGCCATAT 1418
Qy      1391 GCCTTCCCATCTTGGCCTACTACTAA 1418
Db      1419 GCCTTCTCACTGGGCCTACTACTAA 1446

RESULT 6
LOCUS   BC040168
DEFINITION Homo sapiens v-ets erythroblastosis virus E26 oncogene like
          (avian) transcript variant 1, mRNA (CDNA clone MGC:49020
          IMAGE:6052140), complete cds.
ACCESSION BC040168
VERSION   BC040168.1 GI:25304065
KEYWORDS MGC.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2016)
AUTHORS  Krausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
          Altschuler,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D.,
          Hopkins,R.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
          Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
          Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
          Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
          Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
          Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
          McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
          Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
          Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
          Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
          Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
          Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
          Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
          Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
          Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
          Generation and initial analysis of more than 15,000 full-length
          human and mouse cDNA sequences
          Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
          12477932
REFERENCE 2 (bases 1 to 2016)
AUTHORS  Krausberg,R.
TITLE    Direct Submission
JOURNAL  Submitted (22-NOV-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT  Contact: MGC help desk
          Email: csepbe-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Institute for Systems Biology
          http://www.systemsbio.org
          contact: amadan@systemsbiology.org
          Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
          Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 84 Row: 0 Column: 9.
          Location/Qualifiers
          1..2016
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clones="MGC:49020 IMAGE:6052140"
          /tissue_type="Testis, embryonal carcinoma"

FEATURES
Source

```

```

/clone_lib="NIH_MGC_92"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..2016
/genes="ERG"
/db_xref="LocusID:2078"
/db_xref="MIM:165080"
104..1543
/codon_start=1
/product="v-ets erythroblastosis virus E26 oncogene like,
isoform 1"
/protein_id="BAH40168.1"
/db_xref="GI:25304068"
/translation="MASTIKKALSVSDQSLFECAYGTPLAKTMTASSSSDVGQT
SKMSPRVQQDWLSQPARVTIKMECPNSQVGRNRPDECSVAKGKGVGSPDTVGM
NYGSMKEKHPMPNNTERRVTPADPTLWSTDHVQLEWAVKYEGLDVLNILLF
QNIDGKELCKMTKDDFORLTYNADILLSHLYRLTPLPHTLSDVDVKALQNSPRL
MHARNTGAARIFPNTSVYPEATORITTRPDLVPEPPERSAWTGHGHTPOSKAOPS
FSTVFKISDQRPQLDPYQILGFTSRLANPSSGGIQLWQFLLELLSSNSSCITWEG
TNGEFMTDPPDEAVRWGRKSKNNYDKJSLRLLYYDKNIMTKVHKRAYKFD
HGIAALQHPHPESLSLYKPSDLPYMGSYHAHPQOMFVAPHPPALPVTSSSFFAAPN
PYWNSPTGGIYPTNRLPTSHMPSHLGTY"
446..694
/notes="SAM_PNT; Region: Sterile alpha motif (SAM)/Pointed
domain"
/db_xref="CDD:pfam02198"
1031..1282
/notes="Ets; Region: Bts-domain"
/db_xref="CDD:pfam00178"

ORIGIN
Query Match 65.9%; Score 953.4; DB 9; Length 2016;
Best Local Similarity 79.6%; Pred. No. 2.2e-277; Indels 84; Gaps 2;
Matches 1190; Conservative 0; Mismatches 221;
Qy 8 GCGAAGCAATATATTATTAGCAATATTAGCGATCAATATCTTGATCATATGCG 67
Db 49 GAGAAGCAATATATTATTAGCAATATTAGCGATCAATATCTTGATCATATGCG 108
Qy 68 AGGCACTATTAGGAAGCAATATCAGTGTGATGAGTGAAGACCGATCTCTTTGAGTGTGC 127
Db 109 CAGCACTATTAGGAAGCAATATCAGTGTGATGAGTGAAGACCGATCTCTTTGAGTGTGC 168
Qy 128 CTACGGATCGCCACCTTGCACACAGAAATGACAGCCTCTCTTCCAGTGAATATGG 187
Db 169 CTACGGAGCGCACACCTGGCTAGACAGAGATGACCGGTCCTCTCCAGCGACTATGG 228
Qy 188 GCAACATCAAGATGAGCGCGCGGTTCGCCAGCAGGACTGGTTATCAGACGCCCGCGC 247
Db 229 ACAGACTTCCAGATGAGCGCGCGGTTCGCCAGGATTTGGCTGTCTCAACCCCGCAGC 288
Qy 248 CAGAGTTACCATTAAGATGGAGTGAACCCAAACAGGTTAATGGTCAAGGAATTCACC 307
Db 289 CAGGTCACCATCAAAATGGATGTAACTTAGCCAGTGAATGGCTCAAGGAATCTCC 348
Qy 308 TGATCACTCAGCGTGGGCAAGAGGAGAAATGTTAGCAGTTTCAAGCAATTTGGGAT 367
Db 349 TGATCAATGTCAGTGTGGCCAAAGGCGGGAAGATGTGTGGCGCAGCCCGCAGACACCG 408
Qy 368 GAACATGGAAGCTACATGGAGAGAGCAATATTCGGCTCCCAATATGACAAACAATGA 427
Db 409 GAACATGCGCAGCTACATGGAGAGAGCAATATTCGGCTCCCAATATGACAAACAATGA 468
Qy 428 ACGAAGAGTTATTGTGCCAGCAGATCCTACGTTATGGAGCAGACAGACCATGTAGCGAG 487
Db 469 GCGCAGAGTTATCGTGTGCCAGCAGATCCTACGTTATGGAGTACAGACCATGTAGCGAG 528
Qy 488 GCTGAGTGGGCGAGTGAAGAGTATGGTCTTCCAGCTGGGACATCTTGTGTTCCAGAA 547
Db 529 GCTGAGTGGGCGGTGAAGAGTATGGGCTTCCAGACCGTCAACATCTTGTGTTCCAGAA 588
Qy 548 CATTGATGGGAAAGAGTTGTGTAAATGACCAAGATGACTTCCAGAGACTCACGCCGAG 607

```

```

Db 589 CATGATGGGAAGAACTGTGTCAGATGACCAAGAGACATTCAGAGGTCACCCCCAG 648
QY 608 CTATAACCCAGATATCTCTCTGTGTACACCTACATACCTCAGAGAGA----- 654
Db 649 CTACAATCCGACATCTCTCTCTCACATCTCCACTACCTCAGAGAGACTCCTCTCCACA 708
QY 655 ----- 654
Db 709 TTTGACTCAGATGTTGTATGAAGCCTTACAAAATCTCCAGGTTAATGATGCTAG 768
QY 655 -----GAGAGCCACTTTTATTTTCCAAATACATACGTTTACCAGAGCAACGCA 706
Db 769 AAACACAGGGGGTGCAGCTTTTATTTTCCCAATACTTCACTATATCTCAAGCTACGCA 828
QY 707 AAGATATACACACAGGCCAGATTTACCTTATGACAGCGAGGAGATCAGCGTGGAGAG 766
Db 829 AAGAAATTAACAATAGGCCAGATTTACCATATGAGCCCCCAGGAGATCAGCTGGACGG 888
QY 767 TCAGAGCCATCCCA---CTCAGTCAAAAGCTACCAACCATCATCTTCAACAGTGCACAA 823
Db 889 TCAGGGCCACCCAGCCCGAGTCGAAAGCTGCTCAACCATCTCTTCCACAGTGCACAA 948
QY 824 AACAGAGACAGAGCTCTCTAGTTAGATCTTATCAGATTTCTTGACCGACAGCAGCG 883
Db 949 AACTGAACACAGCGTCTCTCAGTTAGATCTCTTATCAGATTTCTTGACCAACAAGTAGCG 1008
QY 884 TCTTCAAAATCCAGGAGTGGCGAGATACAGCTATGGCAGTTCTCTCTGAGCTTCTGTC 943
Db 1009 CTTTGCANAATCCAGCAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCTGAGCTCTGTC 1068
QY 944 GGACAGCTCCAACTCAACTGATCACACTGGAGGGGACAAATGGGAGTTCAAGATGAC 1003
Db 1069 GGACAGCTCCAACTCAACTGATCACACTGGAGGGGACCAACCGGGAGTTCAAGATGAC 1128
QY 1004 AGACCTCATGATGAAGTGGTTCGGCGTTGGGAGAGAGAAAGCAAACTAACATGAACCTA 1063
Db 1129 GGATCCCGACAGAGTGGCCGGCGCTGGGAGAGCGGAGAGCAAAACCAATGAACCTA 1188
QY 1064 TGACAACTCAGCGGTGACCTTCGCTACTACTATGACAAAATATTAATGACTAAAGTCA 1123
Db 1189 CGATAAGCTCAGCGCGCGCTCCGTTACTATGACAAAGAAACATCATGACCAAGGTCCA 1248
QY 1124 TGGTAAACGCTATGCTCAAAATTTGATTTCCAGGATCGCTCAGGCGCTCCAGCCTCA 1183
Db 1249 TGGGAAGCGCTACGCTCAAAATTTGATTTCCAGGATCGCTCAGGCGCTCCAGCCTCA 1308
QY 1184 CCCTCCAGATCATCCATGTACAAATACCCATCAGACCTCCCTACATGAGTTCTCTACCA 1243
Db 1309 CCCCCCGGAGTCACTCTGTACAGTACCCCTCAGACCTCCCGTACATGGCTCTATCA 1368
QY 1244 TGACACACCCGACAGATGAATTTGAGCTCCCATCCCTCCCTGTTGCCCGTAACTC 1303
Db 1369 CGCCACCCACAGAGATGAATTTGAGCTCCCATCCCTCCCTGTTGCCCGTAACTC 1428
QY 1304 ATCCAGCTTTTGTGTCGCCCTTAATCCATACCTGGAATTCACCAACTGAGGACATCTACCC 1363
Db 1429 TTCCAGTTTGTGTCGCCCTTAATCCATACCTGGAATTCACCAACTGAGGATATATACC 1488
QY 1364 CAATACAGGCTGCAGCTGCTCATATGCTTCCATCTTGGACCTACTACTAA 1418
Db 1489 CAACACTAGGCTCCCGACAGCATATGCTTCTCATCTGGGCACTTACTACTAA 1543

RESULT 7
LOCUS AB031088
DEFINITION Rattus norvegicus VESPI4 mRNA for vascular endothelial cell specific protein 14, complete cds.
ACCESSION AB031088
VERSION AB031088.1 GI:15128488
KEYWORDS vascular endothelial cell specific protein 14.
SOURCE Rattus norvegicus (Norway rat)

```

```

ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1808)
Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and
Komurasaki,T.
Identification of VESPI4, a vascular endothelial cell specific
protein
Unpublished
JOURNAL
2 (bases 1 to 1808)
Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and
Komurasaki,T.
Direct Submission
JOURNAL
Submitted (12-AUG-1999) Tomomi Aoki, Taisho Pharmaceutical Co.,
Ltd., Medicinal Research Laboratories, Molecular Biology Lab.,
1-403, Yoshino-cho, Omiya, Saitama 3308530, Japan
(E-mail:sl7704@ccm.taisho.co.jp, Tel:048-663-1111,
Fax:048-652-7254)
FEATURES
Location/Qualifiers
source
1..1808
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/tissue_type="liver"
1..1808
/gene="VESPI4"
63..1430
/gene="VESPI4"
/notes="similar to human erg1"
/codon_start=1
/product="vascular endothelial cell specific protein 14"
/protein_id="BAB62744.1"
/db_xref="GI:15128489"
translation="MASTIKELASVSKDQSLFECAVTPHLAKTMTASSSDYQGT
SKMSRVPQQWLVSOPPARVTIKMNCNQVNSRNPDECSVTGKGVGSPDVTGM
SYGSMKSKHMPNMTNERRVTPADPTLWSTDHVQWLEWAVEYGLLDVDVLLF
QNIDKELCKMTKDDFQRLTPSYADILLHLVRETPLELHSDTDDKALONSPRL
MEARNDLPFRPRRTSTWGHSHPTPQSKAAQPSSTVPKVEDRQPDLPYLIPTLS
SSLANPGSGQQLWQFLLELSDSNCSITWEGTNGBFKMTDDEAVRWGERSKP
NNNYDKSLRLRYYYDKNIMTKVHGKRYAYKFDHGIQAQLQHPPESSLYKYPDLP
YMGSYHTHPQKMFVAPHPALPVTSSSPFATPNPYNWSPSTGGIYPTNRLPASHMPH
LQTY"
Query Match 64.7%; Score 936; DB 10; Length 1808;
Best Local Similarity 80.2%; Pred. No. 4.2e-272;
Matches 1140; Conservative 0; Mismatches 270; Indels 12; Gaps 3;
QY 8 GCGAACGAATAATTATTATTAGCAATATTAGCGATCAATAATCTTGATCAATTATGCG 67
Db 8 GAGAGCAATAATTATTATTAACTTTTAAACGATCAATAAATTTGATTGATGATGCG 67
QY 68 AGCACTATTAAGGAAGCATATATCATGCTGAGTGAAGACCACTCTCTCTTGTGAGTGC 127
Db 68 CAGCACTATTAAGGAGGCTTCTCAGTTGTGAGCAAGACCACTCACTATTTGAGTGC 127
QY 128 CTACGAGTCCGCCACCTTGCAAGACAGAAATGACAGCTCTCTCTTCCAGTGAATATG 187
Db 128 CTACGGAAGCCACACCTGGCTAAGACAGAAATGACCGCATCTCTTCCAGTGAATG 187
QY 188 GCAACATCAAGATGAGCCCGCGGTTCCCGCAGGAGCTGTTATCATCACCCCGCGC 247
Db 188 CCAGACATCAAGATGAGTCCCGAGAGTCCCAACAGGAGCTGCTGTCTCAGCCCCCAG 247
QY 248 CAGAGTTACCATTAAGATGAGTGAACCCCAACAGGTTAATGGTCAAGGAATTCACC 307
Db 248 CAGGTCACATCAAGATGAGTGAACCCCTAGCAGGTTGACCGTTCCAGGACTCAC 307
QY 308 TGATGACTCCAGCGTGGCAAAAGGAGGAATATGTTAGCACTTCAGACAATGTTGGAT 367
Db 308 TGACGAATCAGTGTGACCAAAAGTGGGAAGATGTTGGGCGACCCCTGATCTGTGGGAAT 367
ORIGIN
Query Match 64.7%; Score 936; DB 10; Length 1808;
Best Local Similarity 80.2%; Pred. No. 4.2e-272;
Matches 1140; Conservative 0; Mismatches 270; Indels 12; Gaps 3;
QY 8 GCGAACGAATAATTATTATTAGCAATATTAGCGATCAATAATCTTGATCAATTATGCG 67
Db 8 GAGAGCAATAATTATTATTAACTTTTAAACGATCAATAAATTTGATTGATGATGCG 67
QY 68 AGCACTATTAAGGAAGCATATATCATGCTGAGTGAAGACCACTCTCTCTTGTGAGTGC 127
Db 68 CAGCACTATTAAGGAGGCTTCTCAGTTGTGAGCAAGACCACTCACTATTTGAGTGC 127
QY 128 CTACGAGTCCGCCACCTTGCAAGACAGAAATGACAGCTCTCTCTTCCAGTGAATATG 187
Db 128 CTACGGAAGCCACACCTGGCTAAGACAGAAATGACCGCATCTCTTCCAGTGAATG 187
QY 188 GCAACATCAAGATGAGCCCGCGGTTCCCGCAGGAGCTGTTATCATCACCCCGCGC 247
Db 188 CCAGACATCAAGATGAGTCCCGAGAGTCCCAACAGGAGCTGCTGTCTCAGCCCCCAG 247
QY 248 CAGAGTTACCATTAAGATGAGTGAACCCCAACAGGTTAATGGTCAAGGAATTCACC 307
Db 248 CAGGTCACATCAAGATGAGTGAACCCCTAGCAGGTTGACCGTTCCAGGACTCAC 307
QY 308 TGATGACTCCAGCGTGGCAAAAGGAGGAATATGTTAGCACTTCAGACAATGTTGGAT 367
Db 308 TGACGAATCAGTGTGACCAAAAGTGGGAAGATGTTGGGCGACCCCTGATCTGTGGGAAT 367

```

QY 368 GAACATATGGAGCTACATGGAGAGAGCATATTCGCGCTCCAAATATGACAACTATGA 427
 Db |||||
 QY 368 GAGCTACGCGAGCTACATGGAGGAGAGACATCCCGCCCCCAATATGACCGAGATGA 427
 Db |||||
 QY 428 ACGAAGAGTTATTGTGCGAGCAGATCTCTAGTTATGAGGACACAGCATATGACGGCAGTG 487
 Db |||||
 QY 428 ACGCAGAGTATGCTGCTGCGAGATCTACTCTGTGAGCAGACGACATGCTCGGCGAGTG 487
 Db |||||
 QY 488 GCTGGAGTGGCAGTGAAGGATATGCTCTCCAGAGCTGGACATCTTGTGTTCCAGAA 547
 Db |||||
 QY 488 GCTGGAGTGGCAGTGAAGGATATGCTCTCCAGAGCTGGACATCTTGTGTTCCAGAA 547
 Db |||||
 QY 548 CATTCATGGGAAGAGTTGTGTAATAAGACCAAGATGACTTCCAGAGACTCACGGCGAG 607
 Db |||||
 QY 548 TATTGACGGAAGAGTGTGCAAGATGACAAAGATGACTTCCAGAGGCTCACTCCGAG 607
 Db |||||
 QY 608 CTATTAACGAGATATCTCTGTCACACCTACACTACTCTCAGAGAGA-----GAGGAGCC 662
 Db |||||
 QY 608 CTACAATGCCACATCTTCTCTCACATCTCTCACTACTCTCAGAGAGACTCCCTCTCCACA 667
 Db |||||
 QY 663 ACTTTTATTTTCCAAATACATCAGTTTACCCAGAA---GCAACGCCAAAGATAACAA 718
 Db |||||
 QY 668 TCTGACATCCGACAGCTGTAAGGCTTTACAAACTCTCCAGGTTATGATGCTAG 727
 Db |||||
 QY 719 AAGGCCAGATTTACCTTATGACGAAGCGAGAGATCAGCGTGGACGAGTCACAGCCATCC 778
 Db |||||
 QY 728 AACACAGATTTACCTTATGAGCTCCAGGAGATCAACTGGACCGGCACAGCCACCC 787
 Db |||||
 QY 779 CA---CTCAGTCAAAGTACCCCAACCATCATCTTCAACAGTGCCCAAAACAGAGACCA 835
 Db |||||
 QY 788 CACCCCTCAGTCCAAAGCTGCCAGCCATCTCCCTCCACAGTGCCCAAACTGAAGACCA 847
 Db |||||
 QY 836 CGCTCCTCAGTTAGATCCTTATCAGATCTTGGACCGACAGCAGCGCTTGTGCAAACTCC 895
 Db |||||
 QY 848 GCGTCTCAGTTAGATCCTTACAGATCTTGGACCGACAGTAGTCGCTTGTCTAATCC 907
 Db |||||
 QY 896 AGGAGTGGCGAGATACAGCTATGGAGTTCCTACTGAGCTTCTGTCGACAGCTCCAA 955
 Db |||||
 QY 908 AGGTAGTGGCCAGATCCAGCTGTGGCAGTTCTCTAGAACTCCTGTGTCACAGCTCCAA 967
 Db |||||
 QY 956 CTCCAACTGCATACCTCTGGAGGACCAATATGGGGAGTTCAAGATGACAGACCTGATGA 1015
 Db |||||
 QY 968 CTCCAACTGCATACCTCTGGAGGACCAACCGGGAGTTCAAGATGACAGACCGGATGA 1027
 Db |||||
 QY 1016 AGTGGCTGGCGTTGGGAGAGAGAAAGCAAACTTAACATGAACTATGACAACTCAG 1075
 Db |||||
 QY 1028 GGTGGCTCGGCGCTGGGCGAGCGAAGCAAGCCCAACATGAATGATGACAACTCAG 1087
 Db |||||
 QY 1076 CCGTGCATTTGCTACTACTATGACAAATAATTTATGACTAAAGTTTCATGTTAAACGCTA 1135
 Db |||||
 QY 1088 CCGTGGCCCTCGCTACTACTACGACAAACAAACATGACCAAGGTGACCGGAGCGCTA 1147
 Db |||||
 QY 1136 TGCCTACAAATTTGATTTCCACGGAATGCTCAGGCGCTCCAGCCTCACCTCCAGAACTC 1195
 Db |||||
 QY 1148 TGCCTACAAATTTGATTTCCACGGAATGCTCAGGCGCTCCAGCCTCACCTCCAGAACTC 1207
 Db |||||
 QY 1196 ATCATGTACAAATACCATCAGACCTCCCTCTAGATGATGCTTACCATGACACCCCA 1255
 Db |||||
 QY 1208 GTCCCTGTACAAAGTACCCCTCGACCTGCGTACATGCGGCTCTTATCACACCCACCCCA 1267
 Db |||||
 QY 1256 GAAGATGAACCTTTGTAGCTCCCTCTCCCTCTGCTTTGCGCGCTTACCTCATCCAGCTTTT 1315
 Db |||||
 QY 1268 GAAGATGAACCTTTGTAGCTCCCTCTCCCTCTGCTTTGCGCGCTTACCTCATCCAGCTTTT 1327
 Db |||||
 QY 1316 TGCTGCCCTTAATCATCTAGTGGAAATTCACAACTGGAGGATCTACCGCAATACAGGCT 1375
 Db |||||
 QY 1328 TGCTACCCCGAACCCATCTAGTGGAAATTCGCCAGCTGGGGCATCTACCCGAACTAGGCT 1387
 Db |||||
 QY 1376 GCCAGCTGCTCATATGCTCTCCCTCTTGGCACTTACTACTA 1417
 Db |||||
 QY 1388 CCCAGCCAGCCATATGCTCTCTACCTGGGCACTACTACTA 1429
 Db |||||

RESULT 8
 HUMERG2
 LOCUS Human erg2 gene encoding erg2 protein, complete cds. PRI 08-NOV-1994
 DEFINITION 3166 bp ss-RNA linear
 ACCESSION M17254
 VERSION M17254.1 GI:182186
 KEYWORDS erg 2 protein.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 3166)
 AUTHORS Rao.V.N., Papas,T.S. and Reddy,E.S.
 TITLE erg, a human ets-related gene on chromosome 21: alternative
 splicing, polyadenylation, and translation
 JOURNAL Science 237 (4815), 635-639 (1987)
 MEDLINE 87263429
 PUBMED 3299708
 COMMENT Original source text: Human cell line COLO 320, cdNA to mRNA, clone
 lambda 12.
 FEATURES
 Location/Qualifiers
 1..3166
 /organism="Homo sapiens"
 /mol_type="genomic RNA"
 /db_xref="taxon:9606"
 /map="21q22.3"
 257..1645
 /gene="ERG"
 257..1645
 /gene="ERG"
 /note="erg 2 protein"
 /codon_start=1
 /protein_id="AAA52398.1"
 /db_xref="GI:182187"
 /db_xref="GDB:G00-119-884"
 /translation="MIOVPDPAAHIKKALSVVSEDQSLFECAVGTPLHAKTMTASS
 SSDYQTSKMSPRVQQDWLSQPPARYTIKECNSQVNSNSDECSVAKGKMGV
 SPDTVMGYSTMEKEMPPNINERIVPADPLWSDTHVQWLEWAVKEVGLP
 DNVILLQNI DGKELCKTKDDFORLTSPYNADILLHLHYLRETPLEHLTSDDDVDKA
 LQNSPRLHARNLDLPYPPRRSAWTHGHTFPQSKAAQSPSTVFKETDQSDPDY
 QILGTSRLANPGSQQLQWQLLELLSDNSNSCSITWEGTNGFTKMTDPEAVRW
 GBRKPKMNVKLSRALRYVDKIMTKVHKRYAKFDHGIQAALQHPHPESSLY
 KYPSLPYMGVSHAHPKQMFVAPHPALPVTSSSFFAAPNPFYMSPTGGIYPNTRLP
 TSHMPSHLGTYV"
 repeat_region 291..298
 /note="8-bp repeat sequence"
 repeat_region 509..516
 /note="8-bp repeat sequence"
 polyA_signal 3143..3149
 /note="putative"
 ORIGIN 227 bp upstream of PvuII site; chromosome 21.
 Query Match 63.6%; Score 921; DB 9; Length 3166;
 Best Local Similarity 81.1%; Pred. No. 1.6e-267;
 Matches 1100; Conservative 0; Mismatches 245; Indels 12; Gaps 2;
 QY 74 TATTAGGAAGCATATCAGTGTGAGTGAAGACCAAGTCTTGTGTGAGTGGCTACGG 133
 Db 289 TATCAGGAAGCCTTATCAGTTGTGAGTGAGACCAAGTCTTGTGTGAGTGGCTACGG 348
 QY 134 ATCGCCCCACCTTGCAAGACAGAAATGACAGCTCTCTTCCAGTGAATATGGGCAAC 193
 Db 349 AACGCCACACCTTGAAGACAGATGACCGCTCTCTCCAGCGACTATGGACAGAC 408
 QY 194 ATCAAGATGACCGCGGTTCGCCAGCAGACTGGTTATCAGAGCCCCCGCCAGACT 253
 Db 409 TTCCAAGATGAGCCACGCGTCCCTCAGCAGGATTGGCTGTCTCAACCCCGCCAGGCT 468
 QY 254 TACCATTAAAGTGGAGTGAACCCAAACAGAGTTAATGGGTCAAGGAATTCACCTGATCA 313
 Db 469 CACCATCAATGAATGTAACCTTAGCCAGGTGAATGGCTCAAGGAACCTCTCCTGATGA 528
 QY 314 CTGACGCTGGCAAAAGGAGGAAATGGTTAGCATTGATGGGATGAACTA 373

Db	529	ATCAGTGTGGCCAAAGCGCGGAAGATGGTGGGCAGCCACAGACCGTGTGGGATGAACATA	588
Qy	374	TGGAAGCTACATGGAAGAGAAGCATATTCGGCTTCCAAATATACAAACCAATGAACGAAG	433
Db	589	CGCAGCTACATGGAGGAGAAGCACATGCCACCCCAAACATGACCAACGAGCGCAG	648
Qy	434	AGTTATTGTGGCAGCAGATCCTAGTTATGGAGCAGACCATGTACGGAGTGGGTGGA	493
Db	649	AGTTATCGTGCCAGCAGATCCTACGCTATGGAGTACAGACCATGTGCGCAGTGGGTGGA	708
Qy	494	GTGGCAGTCAAGGAGTATGGTCTTCAGACGTGGACATCTGTGTGTCAGAAACATGA	553
Db	709	GTGGCGGTGAAGAATATGCGCTTCAGAGGTCAACATCTTGTATTCCAGAACATCGA	768
Qy	554	TGGAAAGAGTTGTGTTAAATGACAAAGATGACTTCCAGAGACTCAGCGCGAGCTATAA	613
Db	769	TGGGAAGGAACGTGCAAGATGACCAAGGACGACTTCAGAGGCTCACCCCAAGCTACAA	828
Qy	614	CGCAGATATCCTCCTGTGCACACTACACTACCTCAGAGACA-----GAGGAGCCAC	664
Db	829	CGCGACATCCTTCTCTCAATCTCCACTACTCAGAGAGCTCTCTTCCACATTTGAC	888
Qy	665	TTTTTATTTTCCAAATACATCAGTTTACCAGAGCAACGCAAGAATTAACAAGGCC	724
Db	889	TTCAGATGATGTTGATAAGCCTTACAAACTCTCCACGGTTAATGATCTAGAAACAC	948
Qy	725	AGATTTACCTTATGAGCAACGAGGAGATCAGCGTGGACGAGTCAACGCCATCCCA---C	781
Db	949	AGATTTACCATAATGAGCCCCCAGGAGATCAGCGCTGGACCGGTACGGCCACCCCAAGCC	1008
Qy	782	TCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGCCTCAAAACAGAACACAGCGTCC	841
Db	1009	CCAGTCGAAGCTGCTCAACCATCTCTTCCACAGTCCCAAACTGAAGACCAGCGTCC	1068
Qy	842	TCAGTTAGATCCTTATCAGATTTCTTGAACCGACCGAGCGGTCTGCAAACTCCAGGAG	901
Db	1069	TCAGTTAGATCCTTATCAGATTTCTTGAACCAACAAGTAGCGGCTTGCAATCCAGGCAG	1128
Qy	902	TGGCAGATACAGTATGGCAGTTCTTACTTGGAGTCTCTGTGGACAGCTCCAACTCCAA	961
Db	1129	TGCCAGATCCAGCTTTGGCAGTTCTCTCTTGGAGTCTCTGTGGACAGCTCCAACTCCAG	1188
Qy	962	CTGCATCACCTGGGAGGCGCAAAATGGGAGTTCAAGATGACAGACCTGATGAAGTGC	1021
Db	1189	CTGCATCACCTGGGAGGCGCAACAGCGGAGTTCAAGATGACGATCCCGACAGGTGGC	1248
Qy	1022	TCGGCGTTGGGAGAGGAAAGCAACCTACATGAATGACAACTCAGCGGTGC	1081
Db	1249	CCGGCGTTGGGAGAGCGGAAGCAACCAACATGAATGATGAAGCTCAGCGCGCGC	1308
Qy	1082	ACTTCGCTACTACTATGACAAAAATATATGACTTAAAGTTTCATGGTAAACGCTATGCCTA	1141
Db	1309	CCTCCGTTACTACTATGACAAGAACATCATGACCAAGGTCCATGGGAAGCGTACGCCTA	1368
Qy	1142	CAAAATTTGATTTCCAGGGAATCGCTCAGGCGCTTCAGCGCTCACCTTCCAGAACTATCCAT	1201
Db	1369	CAAGTTCGACTTCCACGGGATCGGCCAGCGGCTTCAGCGCCACGCCCGGAGTCACTCT	1428
Qy	1202	GTCAAAATACCCATCAGACCTCCCTCATATGAGTTCCTACCATGCACACACCCCAAGAT	1261
Db	1429	GTACAGATACCCCTCAGACCTCCCGTACATGGGTCTCTATCACGCCCAACCCACAGAGAT	1488
Qy	1262	GAACTTTGTAGTCCCATCCCTGCTTTGTCGCGTAACTCATPCAGCTTTTTTGTGTC	1321
Db	1489	GAACTTTGTGGCGCCCAACCTCCAGCGCTCCCGTGCATCTTCCAGTTTTTTTGTGTC	1548
Qy	1322	CCCTAATCCATACCTGGAAATTCACCACTGGAGGATCTACCCCAATACAGGCTGCCAGC	1381
Db	1549	CCCAAAACCATCTGGAAATTCACCACTGGGGGTATATACCCCAACTAGGCTCCCAAC	1608
Qy	1382	TGCTCATATGCTTCCCATCTTGGCAGCTACTACTAA	1418

Db 1609 CAGCATTATGCTTCTCATCTCGGCACTTACTACTAA 1645

RESULT 9
XLAJ4126

LOCUS
Xenopus laevis erg gene (erg_E).

DEFINITION
Xenopus laevis erg gene (erg_E).

ACCESSION
AJ224126

VERSION
AJ224126.1 GI:5420047

KEYWORDS
ERG gene; transcription factor.

SOURCE
Xenopus laevis (African clawed frog)

ORGANISM
Xenopus laevis

REFERENCE
1

AUTHORS
Baltzinger M., Mager-Heckel A.M. and Remy P.

TITLE
Xl erg: expression pattern and overexpression during development

JOURNAL
Dev. Dyn. 216 (4-5), 420-433 (1999)

MEDLINE
20099678

PUBMED
10633861

AUTHORS
Baltzinger M.

TITLE
Direct Submission

JOURNAL
Submitted (24-FEB-1998) Baltzinger M., UPR9005, MMDCD, Centre National de la Recherche Scientifique, 15, RUE Rene Descartes, 67084, FRANCE

FEATURES
Location/Qualifiers

source
1. .2012
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="lambda ZAPII-erg E"
/clone_lib="lambda ZAPII"
/dev_stage="stage 3 embryo"
1. .2012
/gene="erg"
225..1595
/gene="erg"
/function="Transcription factor"
/codon_start=1
/protein_id="CA946567.1"
/db_xref="GI:5420048"
/db_xref="GOA:Q9W629"
/db_xref="SPTREMBL:Q9W629"
/translations="MIQTVPDPAHAKAELSVSVESDOSLFEYTGTHLTKTEMTASS
SSDYGTSKMSPRVPOQDMLSQSPSRVTIKWECNPQVNGSRSPDSC1GKSGKMGK
GSDVNGVNGSVYIEEKHISPNWTHERRVVPADPTLWSTDHVRQLEWAIKEYGLP
DNDVLFQNDIGKELCMKEDFRLTPSNADILLSHLYLBERGSAFIIPNSSVYQ
DNQRLSPRODLSPEPSSRAWHNHPAPSKASQPSSTIVKTSDDPRQIDPQILQGT
SSRLANPGSQIQLMGVLELLSDSNNSCITWEGTNGEFPMTDPDFVARWNGRKNK
PNNYTHKALRYYYDKNITMYKHVKRYAKDFDHGIAQALQHPPESTMYKYPSEL
PYNSSYHAHPQKQNFVAPHPALPVTSSSFAPAPNAYWNSPTGSIYENTRLPASHMES
HLGTYT"

gene

CDS

ORIGIN

Query Match 61.5%; Score 889.6; DB 5; Length 1012;
Best Local Similarity 79.7%; Fred. No. 5.1e-258;
Matches 1077; Conservative 0; Mismatches 269; Indels 6; Gaps 2;

QY 74 TATTAAAGAGCATTTACAGTGGTGAAGACACAGTCCCTTGTTGAGTGTGCTACGG 133

Db 257 TATCAAGAGCTTTTCAGTGGTGAAGATCAGTCGTTATTCAGTGACCTACGG 316

QY 134 ATGCCCCCACTTGCAAGACAGAAATGACAGCCTCTCTTCCAGTGAATATGGCAAC 193

Db 317 AACGCCACATCTTACTAAAACGAGATGACCGCATCTTCTCCAGTGAATATGGCAAC 376

QY 194 ATCAAGATGAGCCGCGCTTCCACAGCAGGACTGTTATACAGCCCGCCGACAGT 253

Db 377 CTCGAAATGAGTCCACGTGCTCCACAGACTGGCTCTTCACAGCCTCCATCCAGG 436

QY 254 TACCATTAAAGATGGAGTGTAAACCAACCCAGGTTAATGGGTCAAGGAATTCACCTGATGA 313
Db 437 GACCATCAAGATGGAATGCAACCTAGTCAAGTCAATGGATCAAGAGCTCCCGGATGA 496
QY 314 CTGCAGCGTGGCAAAAGAGGAAATGGTTAGCAGTTTCAGACATGTTGGGATGAACATA 373
Db 497 CTGCAGTATAGGAAAGGAAAGCAAAATGGGTGGAGGTCGGATAATGTTGGAAATGAATTA 556
QY 374 TGGAAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAAACCAATGAACGAAG 433
Db 557 TGGCAGCTACATAGAGAAACACATTTCTCCCAACCAATGACAAACCAATGAGAGGAG 616
QY 434 AGTTATTGTGCGCAGCATCTAGTTATGGAGCAGACCATGTGACGGAGTGGCTGGA 493
Db 617 AGTGATTGTACCTGCTGACCCCTACTTTGTGGAGCAGACATCATGTGACAGAGTGGCTAGA 676
QY 494 GTGGCCAGTGAAGAGATGCTCTTCAGAGCGTGGACATCTGTGTTGTCAGAAACATTTGA 553
Db 677 GTGGCCAAATAAAGAAATACGGCTTCAGATGTCGATGCTCTGTTCCAAACATCGA 736
QY 554 TGGAAAGAGTTGTGTAATGACCAAGATGACTTCCAGAGACTACGCGGAGCTATTA 613
Db 737 CGCAAGGAATATGCAAGATGACCAAGAGATTTCCATCGACTCAGCGCAAGCTACAA 796
QY 614 CGCAGATATCTCTGTCACACCTACACTACCTCAGAGAGAGGAGCCACTTTTATTTT 673
Db 797 TGTGATATCTCTCTCGCATCTTCTACTACTCTCAGAGAAAGAGTGCCTCTTTTATTTT 856
QY 674 TCCAAATACATCAGTTTACCAGAGCAACGCAAGAAATAACCAAGGCGCAGATTACC 733
Db 857 CCCAAATCTATCTCTTATCAAGCGCAACCAAGAAATCCCAAGAGCCAGGCAAGATTATC 916
QY 734 TTATGACAGAGCGAGGAGATCAGCGTGGAGGAGTGCAGAGCATCCCACTCAGTCAAAAGC 793
Db 917 ATATGAGCCCTCAAGAGAGATCTGTTGGCAAAATCA---TCTTGACCAACATCAAAAGC 973
QY 794 TACCAACCATCATCTTCAACAGTGCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 853
Db 974 ATCTCAAC---CATCTACACAGTTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1030
QY 854 TTATCAGATTCTTGGACGAGCAGCGCTCTTGGAAATCCAGGAGTGGGAGATACA 913
Db 1031 TTATCAGATTCTTGGGCAACAGCAGCGCATTTGCAAAATCCAGGAAGTGGGAGATCA 1090
QY 914 GCTATGCGAGTTCTCTATGAGCTTCTGTCGAGCAGCTCCAACTCCAACTGCATCACCTG 973
Db 1091 GCTCTGCAATCTTATGAGCTCTTGGATAGTTCATCCAACTGCATCTACTTG 1150
QY 974 GGAGGCAAAATGGGAGTTCAAGATGACAGACCTGTGATGAAGTGCCTCGCGTGGGG 1033
Db 1151 GGAAGGAACCAATGGAGAAATTAAGATGACCGATCTCTGATGAAGTGGCAAGAGCTTGGG 1210
QY 1034 AGAGAGAAAGCAAACTAACTGAATGACAACTCAGCGTGCCTTCGCTACTA 1093
Db 1211 GGAGAGAAAGCAAACTAACTGAATGACAACTCAGCGTGCCTTCGCTACTA 1270
QY 1094 CTATGACAAAAATATTGACTAAAGTTCTATGTAAGCGTATGCTGCCTACAAATTTGATTT 1153
Db 1271 CTAGTAAATAATATTATGACTAAAGTCCATGGCAAGCGCTATGCTTACAAATTTGATTT 1330
QY 1154 CCAGGAATCGCTCAGCGCTCCAGCTCAGCTCCAGAAATCATCCATGTACAAATACCC 1213
Db 1331 TCATGGGATGCTCAAGCTCTTCAACCTCATCCATGACATCCAGAGATGAATTTGTTG 1390
QY 1214 ATCAGACCTCCCTACATGAGTTCTACATGACACACCCCAAGAGATGAACCTTTGTAGC 1273
Db 1391 CTCAGAACTCCATATATGATGATCATACCATGACATCCAGAGATGAATTTGTTG 1450
QY 1274 TCCCATCCCGCTGTTTGGCCGTAACTGATCCAGCTTTTTCCTGCTCCCTTAATCCATA 1333
Db 1451 CCCCATCCCGCAGCTTTACCTGTGACATCTCAAGTTTCTTTGCGAGCCCTTAATGCATA 1510
QY 1334 CTGGAATTCACCAACTGGAGGCACTACCCCAATACCAAGGCTGCCAGCTGCTCATATGCC 1393

Db 1511 CTGGAATTCACCAACTGGAAGTATTATCCAAATACTCGGCTGCCAGCTAGCCATATGTC 1570
QY 1394 TTCCCATCTTGGACCTACTACTAAGTGGGA 1425
Db 1571 TTCTCATCTTGAACCTACTATTAAATCAAGA 1602
RESULT 10
AB073078 2131 bp mRNA linear ROD 17-OCT-2001
LOCUS Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,
DEFINITION transcript variant 1, complete cds.
ACCESSION AB073078
VERSION AB073078.1 GI:16191716
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.
TITLE Mus musculus Erg mRNA
JOURNAL Published Only in Database (2001)
REFERENCE 2 (bases 1 to 2131)
AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Sushiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
FEATURES
Location/Qualifiers
1..2131
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/chromosome="16"
/clone="Erg-1"
1..2131
/gene="Erg"
148..1536
/gene="Erg"
/note="homolog of human ets-related gene (ERG), transcript variant 1"
/codon_start=1
/protein_id="BAB69948.1"
/translation="MIQTVDPFAAHKEALSIVSEDSQSLFECAYGTPHLAKTEMTPASS
SSDYQTSKSGPRVPQDMLSQAPARVTKMECPNSQVNSRNSPDECSVNGKGVNG
SPDTVMSYSGYMBEKKHVPNNMTTNERVTVPADPTLWSTDHVRQWLEAVKEYGLL
DVDVLLFONIDKELCKMTKDDFORLTPSNADILLSHLYLRETPLPLHSTDVDYDKA
LQNSRLMAEATDLPYEPERRSGTSHSLTPQSKAAQSPSPAVPKTEDORQPDQPY
GILGTSSELANPGSGOILNQFLELLSDSNCSNCTWEGTNGFKVTDPDEAVRRW
GERKSKPNNYDKLSRLRYFYDKNIMTKYHGKRYAYKDFHGIAQALQHPHPESLSL
KPSDLPYNGSYTHAHPQKNFVSPHPALPVTSSFFSPASPNYNSPTGGIYPNTRLP
ASHMPSHLGTY"
ORIGIN
Query Match 60.9%; Score 881.6; DB 10; Length 2131;
Best Local Similarity 79.3%; Pred. No. 1.4e-255;
Matches 1075; Conservative 0; Mismatches 269; Indels 12; Gaps 2;
QY 74 TATTAAGGAAGCATTCAGTGGTGAAGACCAAGTCTCTTCCAGTGAATATGGCTACGG 133
Db 180 TATTAAGGAGGCTTGTGAGTGTGAGCGAGGAGGAGTCACTATTGAGTGTGCTACGG 239
QY 134 ATCGCCCACTTGCAGAGAGAGATGACAGCTCTCTTCCAGTGAATATGGCAAC 193
Db 240 AAGCCACACCTGGTGAAGACAGAGATGACCGCATCTCTTCCAGTGAATATGGCGAGAC 299

Qy	194	ATCAAGATGAGCCGGCGGTTCCCCAGCAGGACTGTATTATCAGACCCCGCCGACGAGT	253
Db	300	ATCCAAGATGAGTCCCAAGAGTCCCTCAGCAGGACTGCGTGTCTCAAGCCCCAGCAGGGT	359
Qy	254	TACATTAAAGATGAGTGTAAACCCAAACCCAGGTTAATGGGTCAAGGAATTCACCTGATGA	313
Db	360	CACCATCAAGATGGAGTGCAAACCTTAGTCAGTGAATGGTTCAGGAACTCACTGATGA	419
Qy	314	CTGCAGCGTGGCAAAAGAGGAGAAATGGTTAGCAGTTCAGACAATGTTGGGATGAACCTA	373
Db	420	GTGCAGTGTGAACAAAGGTGGGAAGATGGTGGGCAGCCCGGATACTGTGGGATGAGCTA	479
Qy	374	TGGAAAGCTACATGAAAGAGAGCATATTCGCGCTCCAAATATGACACCAACCAATGAACGAAG	433
Db	480	CGGAGCTCATATGAGAGAAAGCATGTGCGCCTCCCAATATGACCAAAATGAGCGCAG	539
Qy	434	AGTTATTGTGCCAGCAGATCCTACGTTATGGAGCACAGACATGTACGGCAGTGGCTGGA	493
Db	540	AGTGATCGTCCCTCGAGATCCTACTCTGTGGAGCACAGACATGTCGACAGTGGCTGGA	599
Qy	494	GTGGGCAGTGAAGAGTATGTCCTTCAGACGTGGACATCTGTGTGTTCCAGACACATTGA	553
Db	600	GTGGGCGGTGAAGAAATGCGCTTCCTCGATGTGGAGCTCTTACTATTTCAGAAATTCGA	659
Qy	554	TGGAAAGAGTGTGTAAATGACCAAGATGACTTCAGAGACTCAGCGGAGCTATAA	613
Db	660	TGGAAAGAGCTGTGCAAGATGACAAAGGATGACTTCACGCGGCTCAGCGGAGTACAA	719
Qy	614	CGCAGATATCCTCTGTGCACACCTACACTACTCTCAGAGAG- - - - -GAGAGCCAC	664
Db	720	TGCCGACATTCCTCTCACATCTCCACTACTCCTCAGAGAGACTCCCTTCCACATCTGAC	779
Qy	665	TTTTATTTTCCAAATACATCAGTTTACCCAGAACCAAGCAAGATAACACAGGCC	724
Db	780	TTCGATGACGTTGATAAGGCTTTACAAACTCTCCGGTTAATGCATGCCAGAAACAC	839
Qy	725	AGATTTTACTTATGAGCAAGCGAGAGATCAGCTGTGACGAGTCAAGGCCATC- - -CCAC	781
Db	840	AGATTTTACTTATGAGCTCTCCAGAGATCAGCCTGTGACCGGCCACACGACACCTCAGCCC	899
Qy	782	TCAGTCAAAGCTACCCAAACCATCATCTTCAAACAGTCCCAAACACAGACACGAGCTCC	841
Db	900	TCAGTCAAAGCTGCTCAGCCATCTCCTCTGCAAGTCCCAAACCTGAAACACAGCTCC	959
Qy	842	TCAGTTAGATCCTTATCAGATTCTTGGACGACACGAGCGGTCTTGCAATCCAGGGAG	901
Db	960	TCAGTTAGATCCTTACCAGATCTCTGGGACCGACAGTAGCGGCTTGCTAATCCAGTAG	1019
Qy	902	TGGCGAGTACAGTATGGCAGTTCCTACTGAGGCTTCTGCGACAGCTCCCACTCCAA	961
Db	1020	TGGCCAGATCAGCTGTGGAGTTCCTGCTGCAACTCTGTGACAGAGCTCCCACTCCAA	1079
Qy	962	CTGCATCACCTTGGGAGGSCACAAATGGGGAGTTCAAGATGACAGACCCCTGATGAAGTGGC	1021
Db	1080	CTGCATCACCTTGGGAAGGCCAACCGGGAGTTCAAGATGACAGACCGGACGAGGTGGC	1139
Qy	1022	TCGGGTTGGGAGAGAGGAAGCAACCTTAAATGAATATGACAAATCTCAGCCCTGTC	1081
Db	1140	TCGGCGCTGGGGGAGAGGAAGCAAGCCCAATGAACATGACCAAGTGCACGGGAAGCGTACGCTA	1199
Qy	1082	ACTTCGCTACTACTATGACAAAAATATTATGACTAAAGTTTATGCTAAAAGCTATGCTTA	1141
Db	1200	CCTCGCTACTACTACGACAAAAACATCATGACCAAGTGCACGGGAAGCGTACGCTA	1259
Qy	1142	CAAAATTGATTCCACGGAATGCTCAGGCCCTCAGCCTCACCCCTCCAGAAATCATCAT	1201
Db	1260	CAAGTTTGACTTCCACGGGATTTGCCAGCGCCTCAGCGCCCAACCTCTCTGAGTGGTCCCT	1319
Qy	1202	GTACAAATACCCATCAGACCTCCCTCTACATGAGTTCCTTACCATGCACACCCCAAGAGAT	1261
Db	1320	GTACAAAGTACCCCTCCGACCTGGCCATACATGGGTCTCTATCAGCCCAACCCCAAGAGAT	1379
Qy	1262	GAACCTTTGTAGTCTCCCACTCCCTCTGCTTTCCCGGTAACTTCATCAGCTTTTTTGTCTG	1321

134	QY	ATCGCCCACTCTGCAAGAAGACAGAAATGACAGCGCTCTCTTCCAGTGAATATGGGCAAC	193
240	Db	AACGCCACACTCGCTAAGACAGAGATGACCGCATCTCTTCCAGTCACTATGGCCAGAC	299
194	QY	ATCAAGAATGAGCCCGCGCTTCCACAGCAGGACTGGTTATCAAGCCCGCGCCAGAGT	253
300	Db	ATCCAGATGAGTCCAGAGTCCCTCAGCAGGACTGGCTGTCTCAAGCCCGCAGCAGGGT	359
254	QY	TACCATTAAGATGGAGTGTAAACCCAAACCCAGGTTAATGGGTCAAGGAATTCACCTGATGA	313
360	Db	CACCATCAAGATGGAGTGAACCTAGTCAGGTGAATGGTTCAGGAACATCACTGATGA	419
314	QY	CTGCAGCGTGCCAAAGGAGGGAATAGTTAGCAGTTCAGACATGTTGGGATGAACATA	373
420	Db	GTCCAGTGTGAACAAAGGTGGGAAGATGGTGGGCAGCCCGGATCTACTGTGGGATGAGCTA	479
374	QY	TGGAAGCTCATGGAAGAGAAAGCATATTCGCGCTTCCAAATATGACACACCAATGAACGAAG	433
480	Db	CGCAGCTACATGGAGGAGAAGCATGTGCGCTCTCCAAATATGACCACAATGAGCGCAG	539
434	QY	AGTTATTGCCAGCAGATCCTTAGTTATGGAGCAAGACCATGTACGGCAGTGGCTGGGA	493
540	Db	AGTGATCGTCCCTGCAGATCCTACTCTGTGGAGCACAGACCATGTCCGACAGTGGCTGGA	599
494	QY	GTGGGACGTAAAGAGTATGGTCTTCAGACGTGGACATCTGTGTGTTTCAGACAACATTGA	553
600	Db	GTGGCGGTGAAGAATAAGCCTCTCGATGTGGAGCTTTACTATTTCAGAAATATCGA	659
554	QY	TGGGAAGAGTGTGTAAATAGACCAAGATAGCTTCCAGAGACTCACGCCGAGCTATAA	613
660	Db	TGGGAAGGAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGGCTCAGCCGAGCTACAA	719
614	QY	CGCAGATATCCTCTGTGCACACTACACTACCTACCTCAGAGAGA-----	654
720	Db	TGCGGACATTCTCTCTCACAATCTCCACTACCTCAGAGAGACTCCCTTCCACATCTGAC	779
655	QY	-----	654
780	Db	TTCCGATGAGTTGTATAAGGCTTTACAAAACCTCTCCACGGTTAATGTCATGCCAGAAACAC	839
655	QY	--GAGGAGCACATTTATTTTTCCAAATACATCAGTTTACCGAGACGACGCAAGAAT	712
840	Db	AGGGGTGACGCTTTATTTTCCAAATACTTCAGTATATCCGGAAGCTACGCAAGAAT	899
713	QY	AACAACAAGCCAGATTTACCTTTATGAGCAAGCGGAGAGATCAGCGGTGGAACGAGTCACAG	772
900	Db	TACAACTAGCCACAGATTTACCTTTATGAGCTCCACGAGATCAGCTGGACCGCCACAG	959
773	QY	CGATC---CCACTCAGTCAAAAGTACCCAAACCATCATCTTCAACAGTGCCTCAAAACAGA	829
960	Db	CCACCTCACCCCTCAGTCCAAGCTGCTCAGCCATCTCCCTCTGCACTGCCCAAAACTGA	1019
830	QY	AGACCAAGCTCCTCAGTTAGATCCTTATCAGATTCTGGAACCGACACGACGCGCTCTGC	889
1020	Db	AGACCAAGCTCCTCAGTTAGATCCTTACAGATCTCTGGACCGACAGTAGCCGCTTGC	1079
890	QY	AAATCCAGGAGTGGGAGATACAGCTATGGCAGTTCTCTAGAGCTTCTGTGCGACAG	949
1080	Db	TAATCCAGGTAGTGGCCAGATCCAGCTGTGGCAGTTCTCTCTGCGAACTCTGTGACAGAG	1139
950	QY	CTCAACTCCAACTGCATCACCCTGGGAGGCAACAAATGGGGAGTTCAAGATGACAGACCC	1009
1140	Db	CTCAACTCCAACTGCATCACCCTGGGAGGCAACAAAGGAGTTCAAGATGACAGACCC	1199
1010	QY	TGATGAAGTGGCTCGCGTTGGGAGAGAGGAAAAGCAAACTTAACATGAACATTATGACAA	1069
1200	Db	GGACGAGGTGGCTCGCGCTCGGGGGAGAGGAGAGCAAGCCCAACATGAACATTATGACAA	1259
1070	QY	ACTCAGCGGTGCACTCGCTACTACTATGACAAATAATTTATGACTTAAGCTTCAATGGTAA	1129
1260	Db	GCTCAGCGCGCCCTCGCTACTACTACGACAAAAACATCATGACCAAGGTGACCGGAA	1319
1130	QY	ACGCTATGCTCAAAATTTGATTTCCACGGAATCGCTCAGCGCCTCCAGCCTCACCCCTC	1189

Db	1320	GC	CTACGCCCTACAAGTTTGACCTTCCACGGGATTGCCAGCGCTGTGCAGCGCCACCCCTCC	1379
Qy	1190	AG	ATCATCATCATGTACAAATATACCATCAGACCTCCCTTACATCATGATGTTCTCTACCATGCACA	1249
Db	1380	TG	ATGTCGTCCTGTACAAAGTACCCTCCGACCTGCACATACATGGCTCTTATCAGGCCA	1439
Qy	1250	CC	CCAGAAGATGAACCTTTGTAGCTCCCATCCCCCTGCTTTGCCCGTAACTCATCCAG	1309
Db	1440	CC	CCAGAAGATGAACCTTTGTGTCTCCCACTCCCTCCGCTCTCCAGTCACATCTTCCAG	1499
Qy	1310	CT	TTTTTGCTGCCCCCTAAATCCATACCTGGATTCACCACTGGAGGCATCTACCCCAATAC	1369
Db	1500	TT	CTTTTGCTTCCCGAACCCATACCTGGATTCACCACTGGAGGCATCTACCCCAATAC	1559
Qy	1370	CAG	CTGCCAGCTGCTCATATGCTTCCCATCTTGGCACTTACTACTA	1417
Db	1560	TAG	CTCCAGCAGCATATGCTTCTACCTGGCAGCTTACTACTA	1607
RESULT 12				
AX657230				
LOCUS	AX657230	1291 bp	DNA	linear PAT 22-MAR-2003
DEFINITION	Sequence 3 from Patent WO03000724.			
ACCESSION	AX657230			
VERSION	AX657230.1	GI:29160008		
KEYWORDS	Homo sapiens (human)			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Theria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE 1	Luyten,F., de Bari,C. and Dell'Accio,F.			
AUTHORS	Polynucleotide sequences and vectors useful for the prevention or			
TITLE	treatment of bone- or cartilage-related disorders			
JOURNAL	Patent: WO 0300724-A 3 03-JAN-2003;			
FEATURES	Tigenix N.V. (BE)			
source	Location/Qualifiers			
	1..1291			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			
	146..1228			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAD79712.1"			
	/db_xref="GI:29160009"			
	/db_xref="REMTREMBL:CAD79712"			
	/translation="MYGSDPTVMNYGSYNEEKHMPNNMTINERVIVPADPTLMT			
	DHVRQLEWAVKEYGIDFNILLFQNIIDGKELCKMTKDDFORLTPSYNADILLSHLY			
	LPERGAFFPNTSVYPEATORTPEDLPEPPRESATWGHGHPYQSKAAPSPST			
	VPKTEQRQLDPYQILGTSERLANPGSGQQLQWLFLELLSDSNSCIWEGNG			
	EKFMTDPDEVARKGKRPNNYDKLSKALRYTDKNIMTKVHGSKRYAKPDPFGI			
	AQAQLQHPPESSDIKYPSDLPLPNSYVHAHPQXNFVAPHPPALPVTSSSFFFAAPNPFY			
	NSPTGGIYPNTRLPTSHMPSHLGTY"			
ORIGIN				
Query Match	57.6%; Score 833.6; DB 6; Length 1291;			
Best Local Similarity	84.3%; Pred. No. 4.9e-241;			
Matches 951; Conservative	0; Mismatches 174; Indels 3; Gaps 1;			
Qy	294	TCA	AGGAATTCACCTGATGACTCCAGCGTGGCAAAAGGAGGAAATGFTTACAGTTCA	353
Db	101	TCA	AGGAATTCCTCTGATGAATTCAGTGTGGCCAAAGCGGGAGAGTGTGGGACGCCA	160
Qy	354	GAC	ATGTTGGGATGACATATGAAGCTACATGAAGAGAAGCATATTCGGCTCCAAAT	413
Db	161	GAC	ACGTTGGGATGAATCAGCAGCTACATGAGAGAGAGCATATGCCACCCCAAC	220
Qy	414	AT	GACACCAATGAACGAAGAGTTATTGTGCCAGCAGATCTCTACGTTATGAGCAGAC	473
Db	221	AT	GACACGAACGAGCGCAGATTTCGTGCCAGCAGATCTCTACGTTATGAGCAGAC	280

QY 474 CATGTACGGCAGTGGCTGGAGTGGCGAGTGAAGAGTATGGTCTTCCAGACGTGGACATC 533
 DB 281 CATGTGGCGAGTGGCTGGAGTGGCGGTGAAGAGTATGGCTTCCAGACGTCAACATC 340
 QY 534 TTGTTGTTCCAGAACATGATGAGGAAGAGTGTGTAAATGACCAAGAGTATCCAG 593
 DB 341 TTGTTATTCAGAACATGATGAGGAAGAGTGTGTAAATGACCAAGAGTATCCAG 400
 QY 594 AGACTCAGCGGAGCTATAACGAGATATCTCTGTGCACACCTACACTACCTCAGAGAG 653
 DB 401 AGGCTACACCCAGCTACAGCGGAGATCTCTCTCATCTCCACTACCTCAGAGAG 460
 QY 654 AGAGGAGCAGCTTTATTTTCCAAATACATCATAGTTTACCCAGAGGAGGAAAGATA 713
 DB 461 AGGGGTGACGCTTTATTTTCCAAATACATCATAGTTTACCCAGAGGAGGAAAGATA 520
 QY 714 ACACAGAGGCGAGATTTACCTTTATGACCAAGCAGGAGATGACGCTGGAGAGTACAGC 773
 DB 521 ACACTAGGCGAGATTTACCATATGAGCCCCCAGGAGATCAGCTGGACCGTCAAGC 580
 QY 774 CATCCCA--CTCAGTCAAAAGCTACCAACCATATCTTTCAACAGTGGCCCAAAACAGAA 830
 DB 581 CACCCACGCGCCAGTGCAGAGGTGCTCAACCATCTCTCTCCAGAGTGGCCCAAAACAGAA 640
 QY 831 GACGAGCTCTCAGTTAGATCTTATCAGATTTCTGGACCGGACGAGCGCTTTGCA 890
 DB 641 GACGAGCTCTCAGTTAGATCTTATCAGATTTCTGGACCGGACGAGCGCTTTGCA 700
 QY 891 AATCAGGAGTGGCGAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTCGACAGC 950
 DB 701 AATCAGGAGTGGCGAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTCGACAGC 760
 QY 951 TCCAACTCCAACTGATCACTCTGGAGGAGGACAAATGGGAGTTCAGATGACAGACCT 1010
 DB 761 TCCAACTCCAACTGATCACTCTGGAGGAGGACAAATGGGAGTTCAGATGACAGACCT 820
 QY 1011 GATGAAGTGGCTCGGCGTCTGGGAGAGAGGAAAGCAAACTTAACATGAATATGACAAA 1070
 DB 821 GACGAGTGGCGCGCTCTGGGAGAGAGGAAAGCAAACTTAACATGAATATGACAAA 880
 QY 1071 CTGAGCGGTGACCTTCCTACTACTATGACAAAAATTTATGATTAAGTTCTATGTAA 1130
 DB 881 CTCAGCGCGCGCTTCCTACTACTATGACAAAAATTTATGATTAAGTTCTATGTAA 940
 QY 1131 CGCTATGCTTCAAAATTTGATTTCCACGGAATCGCTCAGGCGCTTCCAGCGCTTCACTCCA 1190
 DB 941 CGCTACGCTTCAAAATTTGATTTCCACGGAATCGCTCAGGCGCTTCCAGCGCTTCACTCCA 1000
 QY 1191 GAATCATCCATGTAACAATACCCATCAGACCTCCCTTACATGATGTTCTTACCATGACAC 1250
 DB 1001 GAGTCATCTCTGTACAAGTACCCCTCAGACCTCCCGTACATGAGGCTCTTATCACGCCAC 1060
 QY 1251 CCCAGAGATGAACTTTGAGTCCCGCATCCCGCTCTTGGCGGTAACCTCATCCAGC 1310
 DB 1061 CCACAGAGATGAACTTTGAGTCCCGCATCCCGCTCTTGGCGGTAACCTCATCCAGC 1120
 QY 1311 TTTTGTGCGCCCTAATCCATATGGAATTCACCAATCGAGGAGTCTACCCCAATACC 1370
 DB 1121 TTTTGTGCGCCCAAAACCAATATGGAATTCACCAATCGAGGAGTCTACCCCAATACC 1180
 QY 1371 AGGCTGCCAGCTGCTATATGCTTCCATCTTGGCAGCTTACTACTAA 1418
 DB 1181 AGGCTGCCAGCTGCTATATGCTTCCATCTTGGCAGCTTACTACTAA 1228

RESULT 13
 LOCUS AB073079 2133 bp mRNA linear ROD 17-OCT-2001
 DEFINITION Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,
 transcript variant 2, complete cds.
 ACCESSION AB073079
 VERSION AB073079.1 GI:16197542
 KEYWORDS

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.
 TITLE Mus musculus Erg mRNA
 JOURNAL Published Only in Database (2001)
 REFERENCE 2
 AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan 1-7-22 Shiohoro-chou,Tsukuba-shi, Ibaraki, Japan 305-8565, Japan (E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
 FEATURES
 source
 1..2133
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /chromosome="16"
 /clone="Erg-2"
 1..2133
 /gene="Erg"
 149..1540
 /gene="Erg"
 /note="homolog of human ets-related gene (ERG), transcript variant 2"
 /codon_start=1
 /protein_id="BAB69949.1"
 /db_xref="GI:16197543"
 /translation="MIQTVPPAAHIKEALSIVSEDSQSFECAYGTHLAKEWTASS
 SSDYGTQKMSPRVQQDULSQAPARTVIRKMECPNQVNGSRNSPDECSVNGKMGV
 SPDYVMSYSGMEERKHPNNNTNRRVIVPADPTLWSDHVRQMLEWAVEKYGLL
 DQVLLQNDIKELCKMDQDFQRLTPSYNADILLSHLYRLRTPPLHLTSDDDVKA
 LQNSPRLHARNTGGAAIFPNTSVYEAQITRPAQAQSPSAVKTEDQRPQLDP
 YQILGSPRLANPGSGQIQWLQELLELLSDSNCSITWGTNGEFKMTDPDEVARR
 WGEKSKPNNVKLSRLRYXDKNTKTVKHKRYAVKDFHGIQAIOHPHPESSL
 VKYPSDLPYMGSYHAHPQKMFVSPHPALPVTSSSPASPNPWNSTGIGIYNTRL
 FASHPMSHLGIYI"
 ORIGIN
 Query Match 57.2%; Score 828; DB 10; Length 2133;
 Best Local Similarity 76.8%; Pred. No.2.5e-239;
 Matches 1046; Conservative 0; Mismatches 295; Indels 21; Gaps 2;
 QY 74 TATTAAGGAAGCATATCAGTGTGAGTGAAGACCAAGTCTCTTGTGAGTGTGCTACGG 133
 DB 181 TATTAAGGAGGCTTGTGAGTGTGAGGAGGAGCCAGTCACTATTGAGTGTGCTACGG 240
 QY 134 ATCGCCCCACCTTGCAGAGACAGAAATCAGACGCTCTCTTCCAGTGAATATGGCAAC 193
 DB 241 AACGCACACCTGGCTAGACAGAGATCCGCATCTCTTCCAGTACTATGGCCAGC 300
 QY 194 ATCAAAGATGAGCCGCGCTTCCCGAGAGATCTGTTATCAGCCCGCCGCGCAGAGT 253
 DB 301 ATCCAAGATGAGTCCCGAGAGTCCCTCAGCAGGAGTGGCTGTCTCAAGCCCGCAGCGGT 360
 QY 254 TACCATTAAAGTGAAGTGTAAACCCAAACAGGTTAATCGGTCAAGGAATTCACCTGATGA 313
 DB 361 CACCATCAAGATGGAATGCACCTTAGTCAAGTGAATGTTCCAGGAATTCACCTGATGA 420
 QY 314 CTGCAAGCTGGCAAGAGGAGGAAATGTTAGCAGTTCAGACAAATGTTGGATGAACTA 373
 DB 421 GTGCAGTGTGAACAAAGGTGGAGAGATGTTGGGAGCAGCCCGGATCTGTGGGATGAGCTA 480
 QY 374 TCGAGAGCTACATGGAGAGAGCATATTCGCGCTCCAAATATGACAACTGAACGAGAG 433
 DB 481 CGGAGCTACATGGAGAGAGAGCATGTGCGCGCTCCCAATATGACCACAATATGAGCGCAG 540

Qy		434	AGTTATTGTGCCACGAGATCCTAGTGTATTGGAGCACAGACAATGTA CGGCAGTGCGCTGGA	493
Db		541	AGTGATCGTCCCTCGAGATCCTACTCTGTGGAGCACAGACAATGTCGCGACAGTGGCTGGA	600
Qy		494	GTGGCGAGTGAAGGAGTATGGTCTTCCAGACGCTGCAATCTGTGTGTTCCAGAACAATGGA	553
Db		601	GTGGCGGTGAAGAATAATGSCCTCCGATGTGACGCTTACTATTTCCAGATATCGA	660
Qy		554	TGGGAAGAAGTTGTAAATGACCAAGAATGACTTCAGAGACTCACGGCAGCTATAA	613
Db		661	TGGGAAGGAGGTGTGCAAGATGACAAAAGGATGATTTCCAGCGGCTCACGGCAGCTACAA	720
Qy		614	CGCAGATATCCTCCTGTGCACACCTACACTACTCTCAGAGAGAGAGGACCACTTTTATTTT	673
Db		721	TGCCGACATTTCTCTCACATCTCCACTACTCTCAGAGAGACTCCCCTTCCACATCTGAC	780
Qy		674	TCCAAATACATCAGTTTTACCACAAGCAA-----CGCAAGAATAAAC	715
Db		781	TTCCGATGACGTTGATAGGCTTTTCAAAACTCTCCACGGTTAATGATGCCAGAAACAC	840
Qy		716	AACAAGGCCAGATTTTACCTTATGACGAAGCAGGAGATCAGCGTGGACGAGTCAACAGCA	775
Db		841	AGGGGTGCAGCTTTATTTTCCCAAATACTTTCAGTATATCCCGAAGCTAGCCAAGAAT	900
Qy		776	TCCCACTCAGTCAAAAGCTACCAACCATCATCTTCAACAGTGGCCCCAAAACAGAGACCA	835
Db		901	TACAACTAGGCC---AGCTGCTAGCCATCTCCCTCTGCACTGGCCCCAAAATGGAAGCA	957
Qy		836	GCGTCTCAGTTAGATCCTTATCAGATTCTTGGACCAACAGCAGCGCTTGTGCAAAATCC	895
Db		958	GCGTCTCAGTTAGATCCTTACCAAGATCCTTGGACCCGACCACTAGCGCCTTGCTAATCC	1017
Qy		896	AGGAGTGGGAGATACAGCTATGGCAATTCCTACTGGAGCTTCTGTGGCAGCTCGAA	955
Db		1018	AGTPTGTGGCAGATCCAGCTGTGGCAGTCTCTGCTCGAACTCCTGTGACAGCTCCAA	1077
Qy		956	CTCCAACTGCATCACCCTGGGAGGCAAAATGGGAGTTCAAGATGACAGACCCCTGATGA	1015
Db		1078	CTCCAACTGCATCACCCTGGGAGGCA CCAACGGGGAGTTCCAGATGACAGACCCGAGCA	1137
Qy		1016	AGTGGCTCGGCTTGGGGAGAGAGGAAAAAGCAAACTTAA CATGAATCATGCAAACTCAG	1075
Db		1138	GGTGGCTCGGCGCTGGGGGAGAGGAGAGCAAGCCCCAACATGAATCATGCAAGCTCAG	1197
Qy		1076	CGGTGCATCTCGCTACTACTATGACAAAAATATTATGACTAAAGTTTATGGTAAACGCTA	1135
Db		1198	CCGGCCCTCGCTACTACTATGACAAAAACATCATGACCAAGGTGACGGGAGCGCTA	1257
Qy		1136	TGCTCAAAATTTGATTTCCACGGAATGCTCTAGCGGCTCCAGGCTCACCTCCAGAAATC	1195
Db		1258	CGCTACAAGTTTGACTTCCACGGGATGGCCAGCGCCTGAGGCCCAACCTCTCTGAGTC	1317
Qy		1196	ATCCATGTACAAATACCATCAGACCTCCCGCTAGATGATGTTCTTACATGACACACCCCA	1255
Db		1318	GTCCCTGTACAAGTATCCCTCCGACCTTGGCATACATGGGCTCTTATCAGGCCACCCCA	1377
Qy		1256	GAAGATGAACTTTGTAGCTCCCATCTCCCTGTCTTTGGCCGTATACCTCATCCAGCTTTT	1315
Db		1378	GAAGATGAACTTTGTCTCCCAACCTCCGCTCTCCCACTCATCTTCTCCAGTTTCTT	1437
Qy		1316	TGCTGCCCTTAATCCATCTGGAATTCACCAACTGGAGGCATCTACCCCAATACAGGCT	1375
Db		1438	TGCTTCCCGNACCCTACTTGGAAATTCACCGACTGGGGCATCTACCGAACACTAGGCT	1497
Qy		1376	GCCAGCTGCTCATATGCTTCCCATCTTGGCACTTACTACTA	1417
Db		1498	CCAGCCAGCATATGCCCTCTACCTGGGCACTACTACTA	1539

RESULT 14				
XLAJ4125	XLAJ4125	1798 bp	mRNA	linear
LOCUS	Xenopus laevis	erg gene (erg A).		
DEFINITION				
				VRT 21-JAN-2000

ACCESSION	AJ224125
VERSION	AJ224125.1 GI:5420045
KEYWORDS	ERG gene; transcription factor.
SOURCE	Xenopus laevis (African clawed frog)
ORGANISM	Xenopus laevis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Anphibia; Batrachia; Anura; Mesobatrachia; Pipoidae, Pipidae;
	Xenopodinae; Xenopus.
REFERENCE	1
AUTHORS	Baltzinger M., Mager-Heckel A.M. and Remy P.
TITLE	Xl erg: expression pattern and overexpression during development pled for a role in endothelial cell differentiation
JOURNAL	Dev. Dyn. 216 (4-5), 420-433 (1999)
MEDLINE	20099678
PUBMED	10633861
REFERENCE	2 (bases 1 to 1798)
AUTHORS	Baltzinger M.
TITLE	Direct Submission
JOURNAL	Submitted (24-FEB-1998) Baltzinger M., UPR9005, MMDCD, Centre National de la Recherche Scientifique, 15, RUE Rene Descartes, 67084, FRANCE
FEATURES	Location/Qualifiers
source	1..1798
	/organism="Xenopus laevis"
	/mol_type="mRNA"
	/db_xref="taxon:8355"
	/clone="lambda ZAPII-erg A"
	/clone_lib="lambda ZAPII"
	/dev_stage="stage 30 embryo"
gene	1..1798
	/gene="erg"
CDS	321..1778
	/genes="erg"
	/function="transcription factor"
	/codon_start=1
	/protein_id="CA846566.1"
	/db_xref="GI:5420046"
	/db_xref="GOA:Q9W700"
	/db_xref="SPTREMBL:Q9W700"
	/translation="MOHMAPIFYEDYSNEALSWSDESLSFECTYGTPLHTKTEMTA SSSDSGVSKMSPRVQPDWLSPQPSRTVKMECPNSOVNGSRSPDDCSIGKGSVM GGSDNMGNYGSIIEKHISPNMNTERRRVIVPADPTLWSTDHYOMLEWAKEVG LPDVLVFQNDIKELCMTKEDPHELTFSYNADILLSHLVHLRETPLPHLTSDDVD KAQNSFRMLHAKTNGASFI PNSSVYQAQMRI PSRQLSYEPERRRSANTWHPPAP SKASQSTTVPTGEDRPOLDPVYLIGTPSSRLANPGSGOI QLUWFELLELLSSDSMN CTIWTEGGEFKMTDPDEVARWGERSKPNNYDKLSRALRYYYDKNIIMTKVHGRRY AYKFDPHGLAQALQHPHPSTWVKYPSELPMSSYHAHPQKNFVAPHPPALPWTSSS FFAAPNAYNSPTGSIYPNTRLPASHMSHLGTYV"
ORIGIN	
Query Match	55.4%; Score 801.8; DB 5; Length 1798;
Best Local Similarity	75.3%; Pred. No. 2.2e-231;
Matches 1079; Conservative	0; Mismatches 267; Indels 87; Gaps 3;
Qy	74 TATTAGGAAGCATTTACAGTGCTGAGTAGAACACCAGTCCTCTGTTGAGTGTGCCTACGG 133
Dd	359 TAGTATGAAGCTCTTTACGTGCTGAGTAGAGATCAGTCCTTATTCCAGTGCACCTACGG 418
Qy	134 ATCCCCCCCCTTGCAAAGACAGAAATGACAGCCTCCTCTTCCAGTGAATATGGCGAAAC 193
Dd	419 AACGCCACACTTTACTATAAACGGAGATGACCGCATCTTCTCCAGTGAATATGGCGAAAC 478
Qy	194 ATCAAGATCAGCCCGCGCTTCCCCAGCAGGACTGGTATTATCACAGCCCGGCCAGAGT 253
Dd	479 CTCGAAATAGTGTCCACGTGTGCTCTCAACAAGACTGGCTCTCACAGCTCTCATCCAGGT 538
Qy	254 TACCATTAAATGGAGTGTAAACCAAACACAGGTTAATGGGTCAAGGAATTCACCTGTATGA 313
Dd	539 GACCATCAAGATGGAATGCAACCTAGTCAAGTCAATGGATCAAGGAGCTCCCGGATGA 598
Qy	314 CTGAGCGTGCRAAAGAGGGAAAATGGTTAGCAGTTACAGCAATGTTGGGATGACTA 373
Dd	599 CTGCAGTATAGRAAAGGAACAAATGGGTGGGGGTCCGATAATGTTGGAATGAATTA 658

Db	510	ACTCCTCTTCCACATTTGACTTCAGATGATGTTGATTAAGCCCTTACAAAACCTCTCCACGG	569
Qy	705	CAAGAATAACAACAGGCCAGATTTTACCTTTATGACCAAGCAGGAGATCAGCGTGGACG	764
Db	570	TTAATGCATCTAGNAACAACAGATTTACCATATGAGCCCCCAGGAGATCAGCCTGGACC	629
Qy	765	AGTCACAGCATCCCA---CTCAGTCAAAAGCTACCCACCATCATCTTCAACAGTGGCC	821
Db	630	GGTCACGGCCACCCACGCCCCAGTGGAGTGGCTCAACCATCTCTCTTCCACAGTGGCC	689
Qy	822	AAAACAGAACACAGCGTCTCAGTTAGATCCTTATCAGATTCTTGACCGACGACGACG	881
Db	690	AAAACTGAAGACACAGCTCTCAGTTAGATCTTATCAGATTCTTGACCAACAAGTAGC	749
Qy	882	CGTCTTGCAAAATCCAGGAGTGGCCAGATACAGCTATGGCAGTTCTTCTGAGCTTCTG	941
Db	750	CGCCTTGCAAAATCCAGGACGTGGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGCTCCTG	809
Qy	942	TCGGACAGCTCCCACTCCAACTGCATCACCCTGGGAGGGCAAAATGGGAGTTCAAGATG	1001
Db	810	TCGGACAGCTCCCACTCCAGCTGCATCACCCTGGGAGGCCAACCGGGAGTTCAAGATG	869
Qy	1002	ACAGACCTTGATGAAGTGGCTCGCGCTTGGGAGAGAGAAAGCAAAACCTAACTGAAC	1061
Db	870	ACGGATCCCGACGAGTGGCCGCGTGGGAGAGCGGAAGCAAAACCAACATGAAC	929
Qy	1062	TATGACAAACTCAGCCGTGGCACTTCGGTACTACTATGACAAAAATATTATGACTAAAGTT	1121
Db	930	TACGATTAAGCTCAGCCCGGCCCTCGTTACTACTATGACGAAGAACATCATGACCAAGGTC	989
Qy	1122	CATGGTAAACGCTATGCTCAAAATTTGATTTCCACGGATCGCTCAGGCCCTCCAGCCT	1181
Db	990	CATGGGAAGCGCTACGCTACAGTTTCGACTTCCACGGATCGGCCAGGCCCTCCAGCCC	1049
Qy	1182	CACCTCCAGAATCATCATGACAAATACCCATCAGACCTCCCTCATCATGAGTCTCTAC	1241
Db	1050	CACCCCCGGAGTCATCTGTACAAGTACCCCTCAGACCTCCCGTACATGGGCTCTCTAT	1109
Qy	1242	CATGCACACCCGAGAAGATGAACTTTGTAGCTCCCCATCCCCCTGCTTTGCCGCTAAC	1301
Db	1110	CAGCCACCCACAGAAGATGAACTTTGTGGCGCCCCACCCCTCCAGCCCTCCCGGTGACA	1169
Qy	1302	TCATCCAGCTTTTTGTGTCGCCCTTAATCCATCTGGAATTCACCAACTGGAGGCATCTAC	1361
Db	1170	TCCTCCAGTTTTTTGTGTCGCCCAACCCCATACTGGHAITTCACCAACTGGGGGTATATAC	1229
Qy	1362	CCCAATACCAAGGTCGACGTGCTCATATGCTTTCCCATCTTGGCACTTACTACTAA	1418
Db	1230	CCCAACTACTAGGTCGCCCAACAGGCCATGCTCTCTCATCTGGGCATTTACTACTA	1286

Search completed: May 27, 2004, 19:25:05
Job time : 5825 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2004, 16:28:01 ; Search time 674 Seconds
(without alignments)
9120.390 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447

Sequence: 1 gaattccgcgaacgaataa.....gaaagaagcgccaagaaa 1447

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1447	100.0	1447	2	AAX26551 DNA encod
2	1356	93.7	1528	2	AAX26552 DNA encod
3	994.4	68.7	1509	8	Abz68766 Nucleotid
4	921	63.6	1389	7	Abx08759 Angiogene
5	921	63.6	3166	7	Abx76295 Lung canc
6	921	63.6	3166	9	Abz75274 Prostata
7	833.6	57.6	1291	8	Abz68767 Nucleotid
8	742.6	51.3	1372	8	Abz68769 Nucleotid
9	698.8	48.3	1219	8	Abz68768 Nucleotid
10	570.2	39.4	1359	8	ADA02689 Mouse Fli
11	570.2	39.4	1359	9	Adb72427 Mouse Fli
12	570.2	39.4	1729	8	ADA02688 Mouse Fli
13	570.2	39.4	1729	9	Adb72426 Mouse Fli
14	563.8	39.0	2938	2	AAX26544 Human Hum
15	562.2	38.9	1359	8	ADA02692 Human Fli
16	562.2	38.9	1359	9	Adb72430 Human Fli
17	562.2	38.9	2957	6	Abv94304 Breast ca
18	562.2	38.9	2957	6	Abx84139 Human cDN
19	562.2	38.9	2957	8	ADA02691 Human Fli
20	562.2	38.9	2957	9	Adb72429 Human Fli
21	544.6	37.6	2954	4	AAX02915 Human she
22	387	26.7	567	4	AAI20932 Probe #10
23	387	26.7	567	4	ABA66003 Human foe

C 24	387	26.7	567	4	AAI46176	Aai46176 Probe #14
C 25	387	26.7	567	4	ABA48124	AbA48124 Human bre
C 26	387	26.7	567	4	ABA33090	AbA33090 Probe #11
C 27	387	26.7	567	4	AAK40160	Aak40160 Human bon
C 28	387	26.7	567	4	AAK14426	Aak14426 Human bra
C 29	387	26.7	567	4	ABS39742	AbS39742 Human liv
C 30	387	26.7	567	5	AAI06643	Aai06643 Probe #66
C 31	387	26.7	567	6	ABS73253	AbS73253 DNA encod
C 32	358.6	24.8	1431	6	ABA68128	AbA68128 Human foe
C 33	354.6	24.5	473	4	ABA50193	Aba50193 Human bre
C 34	354.6	24.5	473	4	ABA50193	Aba50193 Human bre
C 35	354.6	24.5	473	4	ABA35152	AbA35152 Probe #13
C 36	354.6	24.5	473	4	AAK42265	Aak42265 Human bon
C 37	354.6	24.5	473	4	AAK16512	Aak16512 Human bra
C 38	354.6	24.5	473	4	ABS41873	AbS41873 Human liv
C 39	354.6	24.5	473	5	AAI08696	Aai08696 Probe #86
C 40	354.6	24.5	533	3	AAAC01342	AaC01342 Human sec
C 41	325	22.5	533	3	ADA02687	Ada02687 Mouse Fli
C 42	321.4	22.2	72732	8	ADb72425	AdB72425 Mouse Fli
C 43	321.4	22.2	72732	9	ADb72425	AdB72425 Mouse Fli
C 44	320.4	22.1	1890	2	AAQ50662	AaQ50662 Human Hum
C 45	320.4	22.1	73725	8	ADA02690	Ada02690 Human Fli

ALIGNMENTS

RESULT 1
AAX26551
ID AAX26551 standard; DNA; 1447 BP.
XX
AC AAX26551;
XX
DT 14-JUN-1999 (first entry)
XX
DE DNA encoding chicken C-11 protein.
XX
KW Chicken; C-11 protein; cell calcification inhibiting activity;
KW cell calcification inhibiting agent; c-erg protein; arthritis deformans;
KW ossification; spinal column ligament; ss.
XX
OS Gallus sp.
XX
FH Key Location/Qualifiers
FT CDS 63..1418
FT /*tag= a
XX
FN JF11075871-A.
XX
PD 23-MAR-1999.
XX
PF 29-MAY-1998; 98JF-00166076.
XX
PR 18-JUN-1997; 97US-00878177.
PR 20-JUN-1997; 97US-0050297P.
XX
PA (CHUS) CHUGAI PHARM CO LTD.
PA (UYPE-) UNIV PENNSYLVANIA.
XX
DR WPI: 1999-257708/22.
DR P-PSDB; AAY01520.
XX
PT An active protein for inhibiting cell calcification - useful for
measuring the calcification of a cell, for diagnosing arthritis deformans
or ossification of spinal column ligament.
XX
PS Disclosure; Page 7-8; 15pp; Japanese.
XX
CC The present sequence encodes a chicken C-11 protein which has cell
calcification inhibiting activity. The specification also describes a
cell calcification inhibiting agent containing c-erg protein (AAV01521).
CC The proteins are used for measuring the calcification of a cell, for
diagnosing arthritis deformans or ossification of spinal column ligament
CC

PF	08-MAR-2002; 2002WO-EP002605.	QY	554	TGGGAAAGAGTTGTGTATAAATGACCAAGATGACTTCCAGAGACTCAAGCCGAGCTATA	613
XX		Db	579	TGGGAAAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCCGAGTACAA	638
XX	(TIGEX-) TIGENIX NV.	QY	614	CGCAGATATCTCTCTGTGACACCTACCTACCTCAGAGAGAGAGGACCTTTTATTTT	673
XX	Luyten F, De Bari C, Dell'accio F;	Db	639	CGCCGACATCTCTCTCTACATCTCACTCTCAGAGAGGGGTGAGCTTTTATTTT	698
XX	WPI; 2003-221398/21.	QY	674	TCCAAATACATCAGTTTACCCAGAGAGCGCAAGATAACAAAGAGGAGATTTTACC	733
DR	P-PSDB; ABP97695.	Db	699	CCCAAATACCTCAGTATATCTCTGAAAGTACGCAAGAAATTAACAAGTACGAGATTTACC	758
XX	Novel splice variants of the Ets-related gene (ERG) transcription factor	QY	734	TTATGAGCAAGGAGGAGATCAGCTGGACGAGTACAGCCCATCCCA--CTCAGTCAAA	790
PT	, useful for preventing or treating bone-related or cartilage-related	Db	759	ATATGAGCCCCCAGAGAGATCAGCTTGACCGGTACGGCCACCCACGCCCCAGTCGAA	818
XX	diseases or Ets-related gene associated tumor.	QY	791	AGTACCCCAACCATCATCTTTCAAGTGCCTCAAAACAGAGACAGAGCTCTCTCAGTTAGA	850
PS	Claim 9; Fig 4; 60pp; English.	Db	819	AGTGTCTCAACCATCTCTCTCCACAGTGCCTCAAACTGAAGACAGAGCTCTCTCAGTTAGA	878
XX	The present sequence encodes splice variant C-1-1 of human ERG. ERG	QY	851	TCCTTATCAGATCTTGGACCGACAGAGCGCTCTTGCAATCCAGGAGTGGCGAGAT	910
CC	belongs to the Ets family of transcription factors. The specification	Db	879	TCCTTATCAGATCTTGGACCAACAGATAGCGCTTGCAATCCAGGAGTGGCGAGAT	938
CC	describes C-1-1 (characterised by the absence of the 81 bp fragment, and	QY	911	ACAGCTATGGCAGTTCTCTACCTGAGCTTGTGCGACAGCTCCAACTCCAACTGCATCAC	970
CC	the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA	Db	939	CCAGCTTTGGCAGTTCTCTCTGAGCTCTGTCGACAGCTCCAACTCCAGCTGCATCAC	998
CC	sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp	QY	971	CTGGGAGGACAAATATGAGGAGTTCAAGATGACAGACCTGATGAGTGGCTCGGCTTG	1030
CC	fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA	Db	999	CTGGGAGGACCAACGCGGAGTTCAAGATGAGGATCCCGACGAGTGGCCGCGCTG	1058
CC	sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and	QY	1031	GGGAGAGAGAAAGCAACCTTAACATGAATATGACAAACTCAGCCCTGCACTTCGCTA	1090
CC	218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by	Db	1059	GGGAGAGGGAAGAGCAAAACCCACATGAATACGATAGCTCAGCGCGCTCCGTTA	1118
CC	the absence of the 218 bp fragment, and the presence of the 72 bp and 81	QY	1091	CTACTATGACAAATATATGACTAAAGTTTCAATGCTTAAACGCTATGCTACAAATTTGA	1150
CC	bp fragments of the ERG-3 cDNA sequence). The splice variant	Db	1119	CTACTATGACAAAGAACATCATGACCAAGTCCATGGGAGCGCTACGCTCAAGTTTGA	1178
CC	polynucleotides are useful for stabilizing the phenotype of chondrocytes	QY	1151	TTTCCAGGAATCGGTCCAGCGCTCCAGCTCCAGCTCCAGAAATCATCATGACAAATA	1210
CC	or cells with a chondrocytes-like phenotype, for preventing de-	Db	1179	CTTCCAGGGATCGCCAGGCTCCAGCCCCACCCCGGAGTCTCTCTGACAAATA	1238
CC	differentiation of articular chondrocytes during in vitro culture	QY	1211	CGCATCAGACCTCCCTCATAGTGTCTTACATGCAACCCCGCAGAGATGAATTTGT	1270
CC	expansion, for preventing or treating bone-related or cartilage-related	Db	1239	CGCTTCAGACCTCCCGTACATGGGCTCTTATCACGCCCTCCCGTGACATCTTCCAGTTT	1298
CC	diseases or Ets-related gene associated tumor, or as marker of	QY	1271	AGCTCCCATCCCTGCTTTGCGCGTAACTCATCCAGCTTTTGTGCTGCTTAAATCC	1330
CC	progenitor cells of stable chondrocytes. Cells expressing the	Db	1299	GGCGCCCCACCTCCAGCGCTCCCGTGACATCTTCCAGTTTGTGCTGCTTAAATCC	1358
CC	polynucleotide are useful for producing or repairing cartilage tissue in	QY	1331	ATACTGGAATTCACCACTGGAGGATCTTACCCCAATACAGGCTGCCAGCTGCTCATAT	1390
CC	a mammal	Db	1359	ATACTGGAATTCACCACTGGGAGTATATACCCCAACACTAGGCTCCCGCAGCCATAT	1418
XX	Sequence 1509 BP; 389 A; 472 C; 359 G; 289 T; 0 U; 0 Other;	QY	1391	GCCTTCCCATCTTGGCAGCTTACTACTAA	1418
XX	Query Match 68.7%; Score 994.4; DB 8; Length 1509;	Db	1419	GCCTTCTCATCTGGGCACTTACTACTAA	1446
XX	Best Local Similarity 84.1%; Pred. No. 0;	RESULT 4			
XX	Matches 1134; Conservative 0; Mismatches 211; Indels 3; Gaps 1;	ABX08759			
QY	74 TATTAAAGCAATATATCAGTGTGAGTGAAGACGAGCTCTTGTTCAGTGTGCTACGG	ID	ABX08759	standard; cDNA; 1389 BP.	
Db	99 TATCAAGAGACCTTATCAGTGTGAGTGAAGACGAGCTGTGTTGAGTGTGCTACGG	XX	ABX08759;		
QY	134 ATCGCCCACTTGAAGAGAGAAATGACAGCTCTCTTCCAGTGAATATGGGCAAC	AC	ABX08759;		
Db	159 AAGGCCACCTGGCTAAGACAGAGATGACCGCTCTCTCCAGCGACTATGGACAGAC	XX	21-JAN-2003 (first entry)		
QY	194 ATCAAGATGAGCCGCGCTTCCCAGCAGACTGTTATCAAGCCCGCGCCAGAT	DT	Angiogenesis-associated human polynucleotide sequence #21.		
Db	219 TTCCAAGATGAGCCGCGCTTCCCAGCAGACTGTTATCAAGCCCGCGCCAGAT	XX	Human; angiogenesis-associated transcript; angiogenesis;		
QY	254 TACCATTAAGATGAGTGTAAACCCAAACAGGTTAAATGGGTCAAGGAATTCACCTGATGA	DE			
Db	279 CACCATCAATGGAATGTAACTTAGCCAGTGAATGGCTCAAGGAATCTCTCTGATGA	XX			
QY	314 CTGAGCGTGGCAAGAGGGAATGTTAGCAGTTTACAGCAATGTTGGATGAACATA	XX			
Db	339 ATGCACTGTGGCAAGGCGGAATGTTGGCGAGCCAGACCGTTGGATGAACATA	XX			
QY	374 TGGAGCTACATGGAAGAGAGCATTTCCGCTTCCAAATATGACCAATGAACGAAG	XX			
Db	399 CGGAGCTACATGAGAGAGAGCAGCATGCCACCCCAACATGACCGAGAGCGCAG	XX			
QY	434 AGTTATTTGTCAGCAGATCTCTACCTTATGAGGACAGACCATGATGAGCGAGTGGTGA	XX			
Db	459 AGTTATCTGTCAGCAGATCTCTACCTATGAGTACAGACCATGTTGGCGAGTGGCTGA	XX			
QY	494 GTGGCAGTGAAGAGTATGTTTCCAGAGCTGACATCTTGTGTTCCAGAGACATGA	XX			
Db	519 GTGGCGGTGAAGAAATATGGCTTCCAGAGCTCAACATCTGTTATTCAGAACTGA	XX			

KW angiogenesis-associated disease; cancer; cytostatic; gene therapy; gene;
 KW ss.
 XX Homo sapiens.
 OS WO200279492-A2.
 XX 10-OCT-2002.
 PD 14-FEB-2002; 2002WO-US004915.
 XX 14-FEB-2001; 2001US-00784356.
 PR 22-FEB-2001; 2001US-00791390.
 PR 19-APR-2001; 2001US-0285475P.
 PR 03-AUG-2001; 2001US-0310025P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 29-NOV-2001; 2001US-0334244P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA Murray R, Glynn R, Watson SR, Aziz N;
 XX P-PSDB; AEU03476.
 DR WPI; 2003-040681/03.
 XX P-PSDB; AEU03476.
 DR Detecting angiogenesis-associated transcript in a cell for diagnosing and
 PT treating cancer by contacting a sample with a polynucleotide that
 PT exhibits changes in expression level as a function of time in tissue
 PT undergoing angiogenesis.
 XX Example 2; Page 200; 291pp; English.
 PS The present invention relates to methods and compositions for detecting
 CC an angiogenesis-associated transcript in a cell in a patient. The method
 CC involves contacting a biological sample from the patient with a
 CC polynucleotide that selectively hybridises to a sequence at least 80%
 CC identical to any of the angiogenesis-associated human polynucleotide
 CC sequences given in the specification. These angiogenesis-associated
 CC polynucleotide sequences comprise genes that exhibit changes in
 CC expression levels as a function of time in tissue undergoing
 CC angiogenesis. The method and the polynucleotide sequences of the
 CC invention are useful for diagnosing and treating angiogenesis and
 CC angiogenesis-associated diseases e.g. cancer. The polynucleotide
 CC sequences are also useful in the gene therapy of such disorders. The
 CC angiogenesis-associated proteins encoded by the polynucleotide sequences
 CC are useful as a vaccine for therapeutic and prophylactic immunisation.
 CC ABX08739-ABX08853 represent angiogenesis-associated polynucleotide
 CC sequences
 XX Sequence 1389 BP; 359 A; 440 C; 323 G; 267 T; 0 U; 0 Other;
 SQ
 Query Match 63.6%; Score 921; DB 7; Length 1389;
 Best Local Similarity 81.1%; Pred. No. 6.5e-292;
 Matches 1100; Conservative 0; Mismatches 245; Indels 12; Gaps 2;
 74 TATTAAGGAAGCATTATCAGTGGTGAAGACACAGTCTCTTTGAGTGTGCTACGG 133
 33 TATCAAGGAAGCCTATCAGTGTGAGTGAAGACAGTCTGTTTTCAGTGTGCTACGG 92
 134 ATGCCCCACCTTCAAGACAGAAATGACAGCCTCTCTCCAGTGAATATGGCAAC 193
 93 AACGCGACACCTGGCTAAGACAGAGATGACCGCTCTCTCCAGGACTATGGACAGC 152
 194 ATCAAGATGAGCCGCGCTTCCAGCAGAGTGGTTATCAAGCCCCCGGCGCAGGT 253
 153 TTCAAGATGAGCCGCGCTCTCCAGCAGAGTGGTGTCTCAACCCCCAGCAGGTT 212
 254 TACCATTAAGATGGAGTGAACCCAAACAGGTTAATGGGTCAAGGAATTCACCTGATGA 313
 213 CACCATCAAAATGGAATGAACCTTAGCCAGGTGAATGGGTCAAGGAATCTCTGATGA 272
 314 CTGACGCGTGGCAAAAGAGGAGAAATGGTTAGCAGTTCAGACAAATGTGGGATGAAC 373

273 ATGACGTGTGGCCAAAGGGGGAAGATGTGGGCGAGCCAGCAGACCCGTTGGGATGAAC 332
 374 TGGAGCTCATGGAGAGAAAGCATATTCGGCTCCAAATATGACAAACCAATGAACGAAG 433
 333 CGCAGCTCATGGAGAGAAAGCATATTCGGCTCCAAATATGACAAACCAATGAACGAAG 392
 434 AGTTATTTGTCAGCAGATCTTACGTTATGAGACAGACCATGTATGTCGAGTGGCTGGA 493
 393 AGTTATCGTGGCAGCAGATCTTACGTTATGAGACAGACCATGTATGTCGAGTGGCTGGA 452
 494 GTGGCAGTGAAGGATGTGTTCCAGACGTGGACATCTTGTGTTCCAGACATTTGA 553
 453 GTGGCGGTGAAAGAAATATGGGCTTCCAGACGTCAACATCTTGTATTCAGAACATCGA 512
 554 TGGAAAGAGAGTGTGTAAATGACAAAGATGACTTCCAGAGACTCACGCCGAGCTATAA 613
 513 TGGAGAGAACGTGCAAGATGACCAAGGACGATCTCCAGAGGCTCACCCCCAGCTACAA 572
 614 CGCAGATATCTCTCTGTCTACACTTACACTTCTCAGAGAGA-----GAGGAGCCAC 664
 573 CGCCGACATCTCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 632
 665 TTTTATTTTCCAAATACATCATGTTTACCAGAGCAACGCAAGAAATAACCAACAGGCC 724
 633 TTCAAGATGATGTGATTAAGCCTTACAAACTCTCCAGCTTATGTCATGCTAGAACAC 692
 725 AGATTTACCTTATGCAAGAGGAGAGATCAGCTGGAGAGTACAGCCATCCCA---C 791
 693 AGATTTACCTTATGAGCCCCCAGGAGATCAGCTGGACCGCTCAGCGGACCCCCACGCC 752
 782 TCAGTCAAAAGCTTACCAACCATCATCTTCAACAGTGGCCCAAAACAGAGACCCAGCTCC 841
 753 CAGTCAAAAGCTTCTCAACCATCTCTTCCAGTGGCCCAAAACAGAGACCCAGCTCC 812
 842 TCAGTTAGATCTTATCAGATCTTGGACCGACAGCCGCTTGTGAAATCCAGGGAG 901
 813 TCAGTTAGATCTTATCAGATCTTGGACCAACAGTAGCCGCTTGTGAAATCCAGGGAG 872
 902 TGGGAGATACAGCTATGAGTCTCTCTGAGCTTCTGTCGAGCTTCTGTCGAGCTTCTG 961
 873 TGGCAGATCCAGCTTGTGAGCTTCTCTGAGCTTCTGTCGAGCTTCTGTCGAGCTTCTG 932
 962 CTGATCACCTGGGAGGCAAAATGGGAGTTCAGATGACAGACCCCTGATGAAGTGC 1021
 933 CTGATCACCTGGGAGGCAAAATGGGAGTTCAGATGACAGACCCCTGATGAAGTGC 992
 1022 TGGGCTTGGGAGAGAGAAAGCAACCTTAACATGAACTATGACAACTCAGCCGTCG 1081
 993 CCGGCGCTGGGAGAGCGGAGCAACCCCAACATGAACCTAGCTAGCTCAGCCGTCG 1052
 1082 ACTTCGCTACTACTATGACAAATAATTTATGACTAAAGTTCATGTTAAACGCTATGCCTA 1141
 1053 CTTCCGTTACTACTATGACAAAGACATCATGACCAAGGTCCATGGAGCGCTAGCCCTA 1112
 1142 CAAATTTGATTTCCAGGAAATGCTCAGGCCCTCAGGCTCAGCTCAGCTCAGGATCATCAT 1201
 1113 CAAATTTGATTTCCAGGAAATGCTCAGGCCCTCAGGCTCAGCTCAGCTCAGGATCATCAT 1172
 1202 GTACAAATACCCATCAGACCTCCCTTACATGATGTTCTTACCATGACACACCCAGAGAT 1261
 1173 GTACAAATACCCATCAGACCTCCCTTACATGATGTTCTTACCATGACACACCCAGAGAT 1232
 1262 GAATTTGATGTTCCATCCTCCCTTGTGTTGCGGTAACTCATCAGCTTTTGTGCTGC 1321
 1233 GAATTTGTTGGGCGCCCATCCTCCAGCCCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCT 1292
 1322 CCTAATCCATCTGGAATTCACCAATGAGGAGCTTACCCCAATACCAAGCTGCGCAGC 1381
 1293 CCAAAACCATCTGGAATTCACCAATGAGGAGCTTATACCCCAATACCAAGCTGCGCAGC 1352
 1382 TGTCTATGCTTCCATCTTGGCAGCTTACTATA 1418
 1353 CAGCCATATGCTTCT 1389

	D _b	1369	CAGTTTCGACTTTCCACGGGATCGCCAGGCCCTCAGCCCCACCCC CGGAGTCA TCTCT	1428
	O _y	1202	GTCAAATA CCGCATCAGACCTCCCTCATAGAGTTCATGCACACCCCCAGAGAAT	1261
	D _b	1429	GTAACAATACCCCTCAGACCTCCCGTACATGGGTCTCTATCACGCCACCACAGAGAAT	1488
	O _y	1262	GAACTTTGTAGTCCCCATCCCCCTGTTTGCCCGTAACTCATCCAGCTTTTTTGTCTGC	1321
	D _b	1489	GAACTTTGTGGCGCCCCACCCCTCAGACCCCTCCCGTGCATCTTCCAGTTTTTTGTCTGC	1548
	O _y	1322	CCCTAATCCATACTGGAAATTCACCAATGGAGGCATCTACCCCAATACAGGCTGCCAGC	1381
	D _b	1549	CCCAAACCCATCTGGAAATTCACCAATGGGGGTATATACCCCAACACTAGGCTCCCCAC	1608
	O _y	1382	TGTCATATGCTTCCCATCTTGCCACCTACTACTAA	1418
	D _b	1509	CAGCCATATGCTTCTCTCATCTGGSCATCTACTACTAA	1645

```

RESULT 6
ADB75274
ID ADB75274 standard; cDNA; 3166 BP.
XX
XX ADB75274;
XX
XX 04-DEC-2003 (first entry)
XX
XX Prostate cancer marker cDNA.
XX
XX Prostate; cancer; cytostatic; gene therapy; marker; ss.

```

xx The invention relates to newly discovered cancer markers associated with
the cancerous state of prostate cells. Also disclosed is a method of
assessing whether a patient is afflicted with prostate cancer. The method
of the invention involves assessing whether a patient is afflicted with
prostate cancer by comparing the level of expression of a marker in a
patient sample and the normal level of expression of the marker in a
control non-prostate cancer sample, where a significant increase in the
level of expression of the marker in the patient sample and the normal
level indicates that the patient is afflicted with prostate cancer.
Nucleic acids of the invention are useful for diagnosing or treating
prostate cancer, and may be useful in gene therapy. Sequences given in
AB575177-AD575631 represent marker cDNA and proteins. Note: The sequence
data for this patent did not form part of the printed specification, but
was obtained in electronic format directly from WIPO at
cc http://www.wipo.int/pub/publ/pct_sequences.

Db 1249 CCGCGCTGGGAGAGCGGAGAGCAAAACCAATGACATGACGATAAGCTCAGCGCGC 1308
 Qy 1082 ACTTCGCTACTACTATGACAAAATAATTATGACTATAAGTTTCATGGTAAAGCTATGCCTA 1141
 Db 1309 CTTCCGTTACTACTATGACAGAACATCATGACCAAGTTCATGGGAGCGCTACGCTA 1368
 Qy 1142 CAAATTTGATTTCCAGGAATCGCTCAGGCGCTCAGCGCTCAGCTCAGATCATCCAT 1201
 Db 1369 CAAATTGCACTTCCAGCGGATCGCCAGGCGCTCCAGCGCCACCCCGGAGTCACTCT 1428
 Qy 1202 GTACAAATACCATCAGACCTCCCTACATGAGTTCTACATGACACACCCCGAGAGAT 1261
 Db 1429 GTACAGTACCCCTCAGACCTCCGTTACATGGGCTCTCATACGCGCCACCCAGAGAT 1488
 Qy 1262 GAATTTGTAGTCCCATCCCTCCCTCTTTGCCGTAACCTCATTCAGCTTTTGTGTC 1321
 Db 1489 GAATTTGTGCGGCCACCCCTCCAGCGCTCCCTGACATCTTCAGATTTTGTGTC 1548
 Qy 1322 CCCTAATCCATACGGAATTCACCAACTGGAGGATCTCCCAATACCAAGGTCGCAGC 1381
 Db 1549 CCCAAACCATCTGGGAATTCACCAATGGGGTATATACCCCAACTAGGCTCCCCAC 1608
 Qy 1382 TGCTCATATGCTTCCCATCTTGGCACCTACTATA 1418
 Db 1609 CAGCCATATGCTTCTCATCTGGGCACTTACTATA 1645
 RESULT 7
 ABZ68767
 ID ABZ68767 standard; DNA; 1291 BP.
 XX
 AC ABZ68767;
 DT 16-MAY-2003 (first entry)
 DE Nucleotide sequence of human ERG splice variant C-1-2.
 KW Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
 KW C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
 KW cartilage disease; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 Key Location/Qualifiers
 FT 146..1228
 CDS /*tag= a
 FT /product= "ERG splice variant C-1-2"
 FT
 XX
 PN WO2003000724-A2.
 XX
 PD 03-JAN-2003.
 XX
 PF 08-MAR-2002; 2002WO-EP002605.
 XX
 PR 25-JUN-2001; 2001US-0300756P.
 XX
 XX (TIGEX) TIGENIX NV.
 XX
 XX Luyten F, De Bari C, Dell'accio F;
 PI
 XX WPI: 2003-221398/21.
 DR
 DR P-PSDB; ABP97696.
 XX
 XX Novel splice variants of the Ets-related gene (ERG) transcription factor
 PT , useful for preventing or treating bone-related or cartilage-related
 PT diseases or Ets-related gene associated tumor.
 XX
 XX Claim 10; Fig 5; 60pp; English.
 PS
 XX The present sequence encodes splice variant C-1-2 of human ERG. ERG
 CC belongs to the Ets family of transcription factors. The specification
 CC describes C-1-1 (characterised by the absence of the 81 bp fragment, and
 CC the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA

CC sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp
 CC fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA
 CC sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and
 CC 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by
 CC the absence of the 218 bp fragment, and the presence of the 72 bp and 81
 CC bp fragments of the ERG-3 cDNA sequence). The splice variant
 CC polynucleotides are useful for stabilizing the phenotype of chondrocytes
 CC or cells with a chondrocytes-like phenotype, for preventing de-
 CC differentiation of articular chondrocytes during in vitro culture
 CC expansion, for preventing or treating bone-related or cartilage-related
 CC diseases or Ets-related gene associated tumour, or as marker of
 CC progenitor cells of stable chondrocytes. Cells expressing the
 CC polynucleotide are useful for producing or repairing cartilage tissue in
 CC a mammal
 XX
 SQ Sequence 1291 BP; 334 A; 409 C; 302 G; 246 T; 0 U; 0 Other;
 Query Match 57.6%; Score 833.6; DB 8; Length 1291;
 Best Local Similarity 84.3%; Pred. No. 3.9e-263;
 Matches 951; Conservative 0; Mismatches 174; Indels 3; Gaps 1;
 Qy 294 TCAAGGAATTCACCTGATGCTGCGAGCTGCGCAAGGAGGAAATGTTAGCAGTTCA 353
 Db 101 TCAAGGAATTCCTGATGAATGCTGTCGCAAGGAGGAAATGTTAGGAGCCCA 160
 Qy 354 GACAAATGTTGGATGAATGGAAGTACATGGAAGAGAGCAATATTCGCGCTCCAAAT 413
 Db 161 GACACCGTTGGATGAATGCGGAGTACATGGAAGAGAGCAATGCGACCCCAAAAC 220
 Qy 414 ATGACAACCATGAACGAGAGTATTTGTCAGCAGATCTTACGTTATGGAGCAGAC 473
 Db 221 ATGACCAACGAGCGCAGAGTATTCGTCGAGCAGATCTTACGCTATGGAGTACAGAC 280
 Qy 474 CATGTACGGCAGTGGCTGAGTGGGAGTGAAGAGTATGTTCTCCAGAGCGTGACATC 533
 Db 281 CATGTGGGCGAGTGGCTGAGTGGGCGGTGAAGAAATATGGCTTCCAGAGCTCAATC 340
 Qy 534 TTGTTGTTCCAGAACATGATGGGAAGAGTGTGTTAAATGACCAAGAGTACATTCAG 593
 Db 341 TTGTTATTCAGAACATGATGGGAAGAGTGTGTTAAATGACCAAGAGTACATTCAG 400
 Qy 594 AGACTCAGCGCGAGCTATAAGCAGATATCTCTCTGTCAACCTACATCTACCTCAGAGAG 653
 Db 401 AGGCTCAGCGCGAGCTATAAGCAGATATCTCTCTGTCAACCTACATCTACCTCAGAGAG 460
 Qy 654 AGAGGAGCGCATTTTATTTTCCAAATACATCAGTTTACCCAGAGAGCAACGCAAGATA 713
 Db 461 AGGGGTGCGAGCTTTTATTTTCCAAATACATCAGTTTATCTGTAAGCTACGCAAGAAAT 520
 Qy 714 ACAACAGGCGCAGATTTACCTTATGAGCAAGCGAGAGTACGCTGGAACGAGTCAACAGC 773
 Db 521 ACACTAGGCGCAGATTTACCTTATGAGCAAGCGAGAGTACGCTGGAACGAGTCAACAGC 580
 Qy 774 CATCCCA---CTCAGTCAAAAGCTACCCCAACCATCATCTTCAACAGTGCGCCAAACAGAA 830
 Db 581 CACCCCAACGCGCGAGTTCGAAGCTGCTCAACCATCTCTTCCACAGTGGCCAAACCTGAA 640
 Qy 831 GACCGAGCTCTCAGTTAGATCTTATCAGATTCCTTGGACCGGACCGAGCGGCTTTGCA 890
 Db 641 GACCGAGCTCTCAGTTAGATCTTATCAGATTCCTTGGACCGGACCGAGCGGCTTTGCA 700
 Qy 891 AATCCAGGAGTGGGCGAGATACAGTATGCGCAGTTCTCTACTGAGCTTCTGTCGACAGC 950
 Db 701 AATCCAGGAGTGGGCGAGATTCAGATTCAGATTCCTCTGAGCTCTCTGTCGACAGC 760
 Qy 951 TCCAACTCCAACTGATCAGCTGGGAGGCGCAAAATGGGAGTTCAGATGACAGACCTT 1010
 Db 761 TCCAACTCCAGTGCATCATCTACCTGGGAGGCGCAAAATGGGAGTTCAGATGACAGTCC 820
 Qy 1011 GATGAAGTGGCTCGGCTTGGGCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1070
 Db 821 GACGAGTGGCGCGCTGGGAGAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 880

QY 1011 GATGAAGTGGCTCGCGCTTGGGAGAGAGAGAAAGCAAACTAATGAACTATGACAAA 1070
Db |||||
749 GACGAGGTGGCCGCGCTGGGAGAGCGGAGAGCAAACTACATGAACTAGATAAG 808
QY 1071 CTCAGCGTGCACCTCGCTACTACTATGACAAAATATATGACTAAAGTTTCATGGTAAA 1130
Db |||||
809 CTCAGCGCGCGCTCGCTACTACTATGACAAAGAACATCATGACCAAGGTCATGGGAAG 868
QY 1131 CGCTATGCTCAAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCGCTCACCTCCA 1190
Db |||||
869 CGCTACGCTCAAGTTCGACTTCCACGGATCGCCAGCGCTCCAGCGCCACCCCGCG 928
QY 1191 GAATCATCATGTACAAATACCATCAGACTCCCTCATAGATTCCTACCATGACAC 1250
Db |||||
929 GAGTCATCTGTACAAGTACCCCTCAGACCTCCCGTACATGGGCTCTCTATCAGCCAC 988
QY 1251 CCCAGAGATGAATTTTGTAGCTCCCATCCCTGCTTTTGGCCGTAACCTCATCCAGC 1310
Db |||||
989 CCACAGAGATGAATTTTGTGGCGCCACCTCAGCGCTCCCGTGACATCTTCAGT 1048
QY 1311 TTTTGTGCTGCCCCAATTCATCTGGAATTCACAACTGAGGACATCTACCCCAATACC 1370
Db |||||
1049 TTTTGTGCTGCCCCAATTCATCTGGAATTCACAACTGAGGACATCTACCCCAACT 1108
QY 1371 AGGCTGCCAGTCTCATATGCTTCCATCTTGGCAGCTTACTACTAA 1418
Db |||||
1109 AGGCTCCCCACAGCCATATGCTTCTCATCTGGGCACTTACTACTAA 1156
RESULT 10
ADA02689
ID ADA02689 standard; cDNA; 1359 BP.
XX
AC ADA02689;
XX
DT 06-NOV-2003 (first entry)
XX
DE Mouse Flil carcinoma associated coding sequence, SEQ ID NO:1207.
XX
KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ss.
XX
OS Mus sp.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
PS Claim 1; SEQ ID NO 1207; 245pp; English.
XX
CC The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or

CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed murine CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1359 BP; 367 A; 398 C; 331 G; 263 T; 0 U; 0 Other;
Query Match 39.4%; Score 570.2; DB 8; Length 1359;
Best Local Similarity 66.8%; Pred. No. 2.4e-176;
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;
QY 63 ATGCAAGCACTATTAAAGGAAGCATTTATCAGTGGTGAAGAACAGTCCTTTGTTGAG 122
Db |||||
1 1ATGGACGGGACTATTAAAGGAGGCTCTGTCTGTGGTGAAGTGCATCAGTCCCTTTTGTAT 60
QY 123 TGTGCTACGGATCG---CCCCACCTTGCAGAGACAGAAATCAGACGCTCTCTTCCAGT 179
Db |||||
61 TCAGCATACGGAGCGCGACGCCATCTCCCAAGGAGATATGACTGCTTCGGGAGTCTCT 120
QY 180 GAATATGGGCAAAACATCAAGATGAGCCGCGCTTCCCGACGAGCTGGTATTACACAG 239
Db |||||
121 GACTACGGGCGAGCCCGACAAATCAACCCCTGCCACCGCAGGAGTGGATCAACACAG 180
QY 240 CCCCOCGCCAGAGTTACCATTAAGATGGAGTGAACCCAAACAGGTTAATGGTCAAGG 299
Db |||||
181 CC-----AGTGAGAGTCAATGTCAAGCGGGAGTATGACCACATGAATGATCCAGG 231
QY 300 AATTACCTGATGACTGCGAGCGTGGCAAGAGGAGAAATGTTAGCAGTTTCAGACAAT 359
Db |||||
232 GAGTCTCCGGTGGAGTGCAGTGTCAACANTTAACNAGCTGGTGGCGAGGCGAGGCC 291
QY 360 GTTGGGATGAATATGGAAGCTACATGGAAGAGAGCA---TATTCCGCTTCAATATG 416
Db |||||
292 AACCCCATGAATATAATAGTACATGATGATGAGAGACGCGCCCTCTCTCCCAATG 351
QY 417 ACAACCAATGAAGAGAGTATTGTGCCAGCAGATCTTACGTTATGGAGCAGACACCAT 476
Db |||||
352 ACCACCAAGACGGAGAGTCAATGTGCTCGAGACCCACACTGTGGACACAGGACAC 411
QY 477 GTACGGCAGTGGCTGGAGTGGCAGTGAAGAGTATGTTTCCAGACGTGGACATCTTG 536
Db |||||
412 GTTCGACAGTGGCTGGAGTGGCTATAAAGGAATACGGATTGATGGAGATTGACACTTCC 471
QY 537 TTGTTCCAGACATTTGATGGGAAGAGTGTGTAAATACCAAGAGTACCTTCCAGAGA 596
Db |||||
472 TTCTCCAGAACATGATGGAGAGGAAATTTGTAAATGAACAAGAGGAGCTTCTCCGA 531
QY 597 CTCACGCCGAGCTATAACCGCAGATATCTCTCTGTCTCACACTACACTACCTCAGAGAGA 656
Db |||||
532 GCCACCTCCGCTTACACACAGAGTGTGTGTCGACCTCAGTTACCTCAGGGAAG- 590
QY 657 GGAGCCACTTTATTTTCCAAATACATCAGTTTACCCAGAGACGCAAGAAATAACA 716
Db |||||
591 -----TTCACTGCTGGCTTATAACCAACCTCCCATACAGACCGTCTCTCAGCTGAAT 645
QY 717 ACAAGGCCAGATTACCTTTATGAGCAAGCGAGAGATCAGCGTGGACGAGTACACA---GC 773
Db |||||
646 GTCAAGAGAGACCTTCTTATGACTGTCTGAGGAGAGGAGCATGGACATATATGAAC 705
QY 774 CATCCACCTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGGCCCAAAACAGAGAC 833
Db |||||
706 TCTGGCCTCAACAAAAGTCTCTCTCTTGGAGGATCAGACCATGGGCAAGAACACTGAG 765
QY 834 CAGCGTCTCTCAGTTAGATCTTTATCAGATTCTTGGACCGCAGCAGCGCTCTTGCAAT 893

Db 766 CAGCGGCCCCAGCCAGATCCTTATCAGATCTCTGGGGCCAAACGAGCGGCTAGCAAC 825
QY 894 CCAGGAGTGGGCAGATACAGTATGGCAGTTCCTACTGGAGCTTCTGTGCGACAGCTCC 953
Db 826 CTTGGAGTGGGCAGATCCAGCTGTGGAGTTCTCTGGAACTACTCTCGCAGCGGCC 885
QY 954 AACTCCAACTGCATCACCTGGGAGGCGCAAAATGGGAGTTCAAGATGACAGACCTGAT 1013
Db 886 AACGCCAGCTGTATCACCTGGGAGGCGCAACGCGGAGTTCAAAATGACGACCTGAT 945
QY 1014 GAAGTGGCTCGCGTGGGAGAGAGAAAGCAACCTATACGAACTATGACAAATC 1073
Db 946 GAGGTGGCCAGCGCTGGGAGAGCGAGAGCAAGCCCAACATGATATGACAAGCTG 1005
QY 1074 ACCCTGCACCTTCGTACTACTATCACAAAATATATGACTAAAGTTTCATGGTAAACGC 1133
Db 1006 AGCCGGGCCCCCGATCTACTATGACAAACATATATGACAAAGTGCATGGCAAAAG 1065
QY 1134 TATGCTCAAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCTTCACCTCCAGAA 1193
Db 1066 TATGCTCAAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCTTCACCTCCAGAA 1125
QY 1194 TCATCCATGACAAATCCCATCAGACCTCCGCTACATGAGTCTTACCATGACACCCC 1253
Db 1126 ACATCCATGACAAATCCCTCTGATATCTCTACATGSCCTTCTACCATGCCCATCAA 1185
QY 1254 CAGAAAGATGAATTTGTAGTCTCCCATCCCTGCTTTGCGCGTAAACCTCATCCAGCTTT 1313
Db 1186 CAGAAAGTGAATTTGTCCCGTCTCAACCATCTCTCCATGCTGTCACCTCTCCAGCTTC 1245
QY 1314 TTTGCTGCCCCTAATCCATCTGGAATTCACCAACTGGAGCACTACCCCAATACCCAGG 1373
Db 1246 TTTGGAGCAGCATCAATCTGGAACCTCCCGCACTGCTGGGATATATCCAAACCCCGAGT 1305
QY 1374 CTGCCAGCTGCTCATA-----TGCTTCCCATCTTGGCACTACTACTA 1417
Db 1306 GTCCCCCGCATCTTAAACCCACGCTGCTTACACTTAGGCGAGTACTACTA 1358

RESULT 12

ID ADA02688 standard; cdNA; 1729 BP.
XX ADA02688;
AC ADA02688;
XX
DT 06-NOV-2003 (first entry)
XX
DE Mouse Flil carcinoma associated cdNA, SEQ ID NO:1206.
XX
KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ss.
XX
OS Mus sp.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
XX useful for preparing compositions for treating carcinomas.
PS Claim 1; SEQ ID NO 1206; 245pp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03054), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed murine CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1729 BP; 466 A; 484 C; 436 G; 343 T; 0 U; 0 Other;

Query Match 39.4%; Score 570.2; DB 8; Length 1729;
Best Local Similarity 66.8%; Pred. No. 2.8e-176;
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;
QY 63 ATGGCAAGCACTATTAAAGAGCATATCAGTGTGAGTGAAGACCAAGTCTTGTGTGAG 122
Db 245 ATGGACGGGACTATTAAAGAGGCTCTGTGTGTGAGTGAAGTCAAGTCACTCTTTTGTAT 304
QY 123 TGTGCTCTACGATCG--CCACACCTTGCAAGACAGAAATGACAGCTCTCTTCCAGT 179
Db 305 TCAGCATACGGAGCGCAGCCCATCTCCCAAGGAGATATGACTGCTTCGGGAGTCT 364
QY 180 GAATATGGGCAACATCAAGATGAGCCGCGGCTTCCCGAGCAGAGCTGTTATCACAG 239
Db 365 GACTACGGGCGAGCCCGCACAATAATCAACCCCTGCGCCAGCAGAGTGAATCAACAC 424
QY 240 CCCCAGCGCAGAGTTACCATTAAGATGGAGTGTAAACCAACAGCTTAATGGTCAAGG 299
Db 425 CC-----AGTGAGATCATGTCAAGCGGAGTATGACCAATGATGATCCAGG 475
QY 300 AATTCACTGATGACTGCGCGGTGGCAAAAGGAGGAAATGTTAGTAGTTCAGACAT 359
Db 476 GAGTCTCCGCTGACTGCACTGTCAAGAAATGTAAACAGCTGTGTGGGCGGAGCGAAGCC 535
QY 360 GTTGGGATGAACCTATGGAAGCTACATGGAAGAGAGCA--TATTCCGCTCCAAATATG 416
Db 536 AACCCATGAACCTATTAATAGCTTACATGGATGAGAGAACGGCCCCCTCTCCACATG 595
QY 417 ACAACCAATGAACGAAGAGTTATTGTGCCAGCAGATCTTACGTTATGGAGCAGACCAT 476
Db 596 ACCACCAACGAACGGAGAGTCAATTTGTCTGCAAGACCCCACTGTGTGACACAGGAGCAC 655
QY 477 GTACGGAGTGGCTGAGTGGCGCAGTGAAGAGTATGTTCTCCACAGCTGGGACATCTTG 536
Db 656 GTTCGACAGTGGCTGAGTGGCGCTATAAAGGAATACGGATATGAGAGATGACATCTCC 715
QY 537 TTGTTCCAGAACTTGTATGGGAAAGAGTTGTGTAAATGACCAAGATGACTTCCAGAGA 596
Db 716 TTCTTCCAGAACATGATGGCAGGAAATTTGTGTAAATGACCAAGGAGGACTTCTCCGA 775
QY 597 CTCACGCGGAGCTATAACGAGATATCTCTCTGTGTCACTTACCTACCTCAGAGAGA 656
Db 776 GCCACCTCCGCTTACCAACACAGAAAGTGTGTGTGTCGACCTCAGTTTACCTCAGGAAAG- 834
QY 657 GGAGGCCACTTTTATTTTTCACAAATACATGTTTACCCAGAGCAACCGCAAGAAATAACA 716
Db 835 -----TTCAGTGTGGCTTATAACACAACTCCCATAGAGACCACTCTCACAGTGAAT 889

QY 717 ACAGGGCAGATTTACCTTATGACGACGAGGAGATCAGCGTGGAGGATCACA---GC 773
 Db 890 GTCAGGAGAGCCCTTCTATGACTCTCTCAGGAGAGGAGATGGAACATAATATGAAC 949
 QY 774 CATCCACTAGTCAAAAGCTACCAACCATCATCTTCAACAGTSCCCAAACAGAGAC 833
 Db 950 TCTGGCCTCAACAAAGTCTCTCTCTGGAGATCAGAGCCATGGCAAGACACTGAG 1009
 QY 834 CAGCGTCCCTCAGTTAGATCCCTTATCAGATTTCTGGACCGCAGCAGCGCTCTGCAAT 893
 Db 1010 CAGCGGCCCGCAGCAGATCCCTTATCAGATCTCTGGGCCAACACAGCGCCGCTAGCAAC 1069
 QY 894 CCAGGAGTGGGCGATACAGCTATGGCAGTTCCTTACTGGAGCTTCTGTGGACAGCTCC 953
 Db 1070 CCTGGAGTGGGCGATCAGCTGGCGAGTTCTCTGGAACTACTGTCGACAGCGCC 1129
 QY 954 AACTCCACTGATCAGTCTGGAGGACAAATGGGAGTTCAGATGACAGACCTTGAT 1013
 Db 1130 AACGCCAGTGTATCAGCTGGAGGGGACCAACGGGAGTTCAAAATGACGGACCTGAT 1189
 QY 1014 GAAAGTGGCTCGCGTGGGAGAGAGGAAAGCAAACTTAACATGAATATGACAAACTC 1073
 Db 1190 GAGGTGGCCAGCGCTGGGAGAGCGGAGAGCAAGCCCAACATGAATTATGACAGCTG 1249
 QY 1074 AGCGGTGACCTTGGTACTATATGACAAAAATATATGACTAAAGTTCATGTAAACGC 1133
 Db 1250 AGCGGGGCCCTCCGATACTATGACAAAAACATTTATGACAAAGTGCATGGCAAGG 1309
 QY 1134 TATGCTACAAATTTGATTTCCAGGAATGCTCAGGCGCTCCAGCGCTCAGCCATCA 1193
 Db 1310 TATGCTACAAAGTTGATTTCAATGGCATTTGCCAGGCCCTGAGCCACATCCAAAGAG 1369
 QY 1194 TCATCCATGTACAAATCCCATACAGACCTCCCTTACATGATGATTCCTACCATGCAACCCC 1253
 Db 1370 ACATCCATGTACAAATCCCTCTGATATCTCTACATGCTTCTTCTACCTGCGCATCAA 1429
 QY 1254 CAGAGATGAATTTGATGCTAGCTCCCATCCCTCTGTTTGGCCGTAACCTCATCCAGCTTT 1313
 Db 1430 CAGAAAGTGAATTTGCTCCGTTCTCACCATCTCTCATGCTGTCATCTCTCCAGCTTC 1489
 QY 1314 TTTGCTGCCCTTAATCCATCTGGAATTCACCAACTGGAGGCATCTACCCCAATPACAGG 1373
 Db 1490 TTTGGAGCAGCATCAATATCTGACCTCCCGCTCCCGCTGCTGGATCTATCCAAACCCAGT 1549
 QY 1374 CTCGACAGTGTCTATA-----TGCTTCCCATCTTGGCAGCTACTACTA 1417
 Db 1550 GTCCCCCGCCATCCTTAACACCCACGTCGCTTTCACACTTAGGCGAGCTACTACTA 1602

RESULT 13
 ADB72426
 ID ADB72426 standard; mRNA; 1729 BP.
 AC AC ADB72426;
 XX DT 04-DEC-2003 (first entry)
 XX DE Mouse Flil mRNA.
 XX KW mouse; ss; cytosstatic; gene therapy; vaccine; carcinoma; lymphomas;
 XX LW cancer; neoplasm; adenocarcinoma; sarcoma.
 XX OS Mus sp.
 XX PN WO2003008583-A2.
 XX PD 30-JAN-2003.
 XX PF 26-DEC-2001; 2001WO-US051291.
 XX PR 02-MAR-2001; 2001US-00798586.
 XX PR 23-OCT-2001; 2001US-00004113.

PR 08-NOV-2001; 2001US-00052482.
 PR 30-NOV-2001; 2001US-00997722.
 PR 20-DEC-2001; 2001US-00034650.
 XX (SAGR-) SAGRES DISCOVERY.
 PA Morris DW, Engelhard EK;
 PI WPI; 2003-239337/23.
 XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.
 XX Claim 1; SEQ ID NO 254; 2304pp; English.
 PS The invention relates to a novel recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the 660 sequences fully defined
 CC in the specification. A polynucleotide of the invention has cytostatic
 CC activity, and may have a use in gene therapy, or in a vaccine. The
 CC recombinant nucleic acids and polypeptides are useful for treating
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
 CC sarcomas. The present sequence represents a mouse mRNA of the invention.
 XX SQ Sequence 1729 BP; 466 A; 484 C; 436 G; 343 T; 0 U; 0 Other;

Query Match 39.4%; Score 570.2; DB 9; Length 1729;
 Best Local Similarity 66.8%; Pred. No. 2.8e-176;
 Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;
 QY 63 ATGGCAAGCATTATTAAAGGAGCATTTATCAGTGGTGGTGAAGACAGAGCTCTTTTGGAG 122
 Db 245 ATGGCAGGACATTATTAAAGGAGCTCTGCTGTGGTGGTGAAGACAGATCAGTCCCTTTTGTAT 304
 QY 123 TGTGCTACGGATCG---CCCCACCTTGAAGAAGACAGAAATGACAGCTCTCTTCCAGT 179
 Db 305 TCAGCATACGGAGCGGCGAGCCCATCTCCCAAGGACAGATATGACTGTTCGGGGAGTCT 364
 QY 180 GAATATGGGCAACATCAAGATGAGCCGCGCTTCCACAGAGCTGGTTATCACAG 239
 Db 365 GACTACGGCAGCCCAAAATCAACCCCTGCGCACCGCAGAGGTGGATCAACAG 424
 QY 240 CCCCCGGCAGAGTTACCATTAAGATGGAGTGTAAACCAACACAGTTAATGGGTCAAGG 299
 Db 425 CC-----AGTGAGAGTCAATGTCAAGCGGAGTATGACCATGAATGGATCCAGG 475
 QY 300 AATTACCTGATGACTGCGAGCGTGCBAAGAGGAGGAAATGTTAGCAGTTTCAGACAAT 359
 Db 476 GAGTCTCCGGTGGAGTGCAGTGTCAAGCAATTAACAGCTGGTGGCGGAGCGGAGGCC 535
 QY 360 GTTGGGATGAACCTATGGAAGCTACATGGAAGAGAGCA---TATTCGGCTCCCAATATG 416
 Db 536 AACCCCATGAACATAATAGCTACATGGATGAGAAGACGGCCCTCTCTCCCAACATG 595
 QY 417 ACACCAATGAACGAGAGTATTGTGCGCAGAGATCTTACGTTATGGAGCAGACCAT 476
 Db 596 ACCACCAAGACGGAGAGTCAATGTGCTGAGACCCCACTGTGGACAGAGGAC 655
 QY 477 GTACGGCAGTGGTGGAGTGGGCGAGTGAAGAGTATGGTCTTCAGAGCTGGACATCTTG 536
 Db 656 GTTGGACAGTGGCTGGAGTGGCTATAAGGAATACGGATTGATGGAGATTGACACTTC 715
 QY 537 TTGTTCCGAAACATTGATGGGAAAGAGTTGTATAAATGACCAAGAGTATGATCCGAGA 596
 Db 716 TTCTTCCGAAACATTGATGGGAAAGAGTTGTATAAATGACCAAGAGAGTCTCTCCGA 775
 QY 597 CTCACGCCGAGCTATAACGACAGATATCTCTCTGTGCACACTACCTACCTCAGAGAGA 656
 Db 776 GCCACCTCCGCTACACACAGAGTGTGTGTCGACCTCAGTTACCTCAGGGAAG- 834
 QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCGAGAGCAGCAAGAGATAACA 716
 Db 835 -----TTCACTGTGCGCTTATAACCACTCCCATACAGACCACTCTCTCAGACTGAAT 889

QY 717 ACAAGGCCAGATTACCTTATGAGCAAGCGAGAGATCAGCGTGGACGAGTCAAC---GC 773
 Db 890 GTCAGGAGAGACCCCTTCTTATGACTCTGTGAGGAGAGGAGCATGGAACAATAATATGAAC 949
 QY 774 CATCCACTCAGTCAAAAGCTACCAACCATCATCTTCAAGTGCACCAACAGAGAC 833
 Db 950 TCTGGCCTCAACAAGACCTCTCTCTGGAGATACAGACCATGGGCAAGAACACTGAG 1009
 QY 834 CAGCGTCTCAGTTAGATCTTATCAGATTCTTGGACCGAGCAGCCGCTTTGCAAAAT 893
 Db 1010 CAGCGGCCCGCAGATCTTATCAGATCTGTGGGCCAACACAGCAGCCGCTTAGCAAC 1069
 QY 894 CAGGAGTGGGAGATACACTATGGCAGTCTCTACTGAGCTTCTGTGGCAGCTCC 953
 Db 1070 CTTGGAGTGGGAGATCCAGCTGTGGCAGTTCTCTCTGGAACTACTGTCCGACAGCC 1129
 QY 954 AACTCCAACTGATCACCCTGGGAGGACAAATGGGAGTTCAAGATGACAGACCTGAT 1013
 Db 1130 AACGCCAGCTGTATCACCCTGGAGGGGACCAACGGGAGTTCAAAATGACGACCTGAT 1189
 QY 1014 GAAGTGGCTCGGCTTGGGAGAGAGAAAGCAAACTTAACATGAACATATGACAACTC 1073
 Db 1190 GAGGTGGCCAGCGCTGGGAGAGCGGAAGCAAGCCCAACATGAATATGACAACTG 1249
 QY 1074 AGCGTGACATCTGCTACTATGACAAATAATATGACTAAAGTTCAAGTTGTAACGC 1133
 Db 1250 AGCGGGCCCTCCGATCTACTATGACAAACATTAATGACAAAGTGCATGGCAAGG 1309
 QY 1134 TATGCTCAAAATTTGATTTCCAGGAATCGCTCAGGCGCTCCAGCTTACCTCCAGAA 1193
 Db 1310 TATGCTCAAAATTTGATTTCCAGGAATCGCTCAGGCGCTCCAGCTTACCTCCAGAA 1369
 QY 1194 TCATCCATGACAAATCCCATCAGACCTCCCTACATGAGTTCCTACATGACACACCC 1253
 Db 1370 ACATCCATGACAAATCCCATCAGACCTCCCTACATGAGTTCCTACATGACACACCC 1429
 QY 1254 CAGAGATGAATTTGTAGCTCCCATCCCTGCTTTGCCCGTAACTCATCAGCTTT 1313
 Db 1430 CAGAGATGAATTTGTAGCTCCCATCCCTGCTTTGCCCGTAACTCATCAGCTTT 1489
 QY 1314 TTTGCTGCCCTAATCCATCTGGAATTCACCACTGAGGACATCTACCCCAATACGAG 1373
 Db 1490 TTTGAGCAGCATCAATCAATCTGAGCTCCCGCACTGCTGGATCTATCCAAACCCAGT 1549
 QY 1374 CTGCCAGCTGCTCAFA-----TGCCCTTCCCATCTTGGCAGCTACTACTA 1417
 Db 1550 GTCCCCCGGCATCTTAACACCCACGCTGCTTACACTTAGGACGCTACTACTA 1602

RESULT 14
 AAQ50644
 ID AAQ50644 standard; cDNA; 2938 BP.
 AC AAQ50644;
 XX
 XX
 DT 25-MAR-2003 (revised)
 DT 26-MAY-1994 (first entry)
 XX
 DE Human Hum-Flu-1 gene clone BM025.
 DE
 XX chromosomal translocation; chimeric; chimeraic; Ewing sarcoma; Ews gene;
 KW malignant melanoma; hum-flu-1;
 KW primitive peripheral neuroectodermal tumour; human chromosome 11;
 KW human chromosome 22; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 143..1501
 FT /tag= a
 FT /product= "HUM-FLI-1"
 FT 2908..2913
 FT poly_a_signal
 FT /*tag= b

XX WO9323549-A2.
 PN 25-NOV-1993.
 XX 19-MAY-1993; 93WO-FR000494.
 PF 20-MAY-1992; 92FR-00006123.
 PR (CNRS) CNRS CENT NAT RECH SCI.
 PA Aurias A, Delattre O, Desmaziere C, Melot T, Peter M, Plougastel B;
 PI Thomas G, Zucman J;
 XX WPI; 1993-386590/48.
 DR P-PSDB; AAR44556.
 XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence
 PT involved in chromosomal translocation, also derived mRNA, probes, fusion
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
 XX Disclosure; Fig 7; 123pp; French.
 XX The probe 11R1 was used to screen a human marrow cDNA library (Clontech
 CC cat.# HU1058). The clone BM025 was identified and sequenced. It
 CC represents the entire coding region together with 5' and 3'-UTRs of the
 CC Hum-Flu-1 gene. (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 2938 BP; 847 A; 692 C; 675 G; 724 T; 0 U; 0 Other;
 Query Match 39.0%; Score 563.8; DB 2; Length 2938;
 Best Local Similarity 56.5%; Pred. NO. 58-174;
 Matches 913; Conservative 0; Mismatches 427; Indels 33; Gaps 6;
 QY 63 ATGGCAAGCAGCTATTAAAGGAGCATTTATCAGTGTGAGTGAAGACAGCTCTTGTGTGAG 122
 Db 143 ATGGAGCGGAGCTATTAAAGGAGCTCTGTGCGTGTGAGGAGCAGCAGCAGCTCTTGTGAC 202
 QY 123 TGTGCTTACGATCG---CCCCACCTTGAAGACAGAAATGACAGCTCTTCTTCCAGT 179
 Db 203 TCAGCTACGGAGCGCAGCCCATCTCCCAAGCGCAGCATGACTGCTCGGGAGTCTCT 262
 QY 180 GAATATGGGCAACATCAAGATGAGCCCGGTTCCCGCAGCAGCTGTTTATCAGAG 239
 Db 263 GACTACGGGAGCCCAAGATCAACCCCTCCACAGCAGCAGGAGTGATCAATCAG 322
 QY 240 CCCCCGCGCAGCTTACCATTAAGATGAGTGAACCCAAACCCAGCTTAAATGGTCAAGG 299
 Db 323 CC-----AGTGAGGTCAACGTCAAGCGGAGTATGACCATGAATGATCCAGG 373
 QY 300 AATTCACTGATGACTGCGAGCGTGGCAAGAGGAGAAATGTTAGCAGTTCAGACAT 359
 Db 374 GAGTCTCGGTGACTGCGAGCTTAGCAAAATGCGCAAGCTGTGGCGGAGCGAGTCC 433
 QY 360 GTTGGGATGAACCTATGGAAGTACATGGAAGAGAGAGCAT---ATTCCGCTCCAAATATG 416
 Db 434 AACCCATGAACCTATGGAAGTACATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493
 QY 417 ACAACCAATGAACGAGAGAGTATTGTGCCAGCAGATCTTACGTTATGGAGCAGACCAT 476
 Db 494 ACCACCAACGAGAGAGAGAGTATCGTCCCGCAGAGCCCACTGTGGACACAGAGAGCAT 553
 QY 477 GTAGCAGTGGCTGAGTGGGAGTGGGAGTGGTCTTCCAGACGTTGGAGCATCTTG 536
 Db 554 GTAGGCAATGGCTGGAGTGGGCCATTAAGAGAGTATAGCTGTATGGAGATCGACATCC 613
 QY 537 TTGTTCCAGAACATTTGATGGGAAAGAGTTGTGTAAATATGACAAAGATGATCTTCCAGAG 596
 Db 614 TTTTCCAGAACATTTGATGGGAAAGAGTGTGTAAATATGAAACAGAGAGAGAGTCTTCCCGC 673
 QY 597 CTCACGCCGAGCTATTAACGAGATATCTTCTCTGTCACCTTACACTACCTCAGAGAGAG 656
 Db 674 GCCACCAACCTCTTACACACAGGAGTCTGTTGTACACCTTACCTTACCTCAGGGAAG- 732

```

QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAGCAACGCAAGAAATAACA 716
Db 733 -----TTCACTGCTGGCCATATAATACAACTCCACACCGACCAATCTCTCAGATTGAGT 787
QY 717 ACAGGCCAGATTTTACCTTATGAGCAGCAGGAGAGATCAGGTGGACAGTTCACAGCCAT 776
Db 788 GTCAAGAGAGACCCCTTCTTATGACTCAGTCAGACAGAGAGAGAGATGGGCAATACATGAAT 847
QY 777 CC---CACTCAGTCAAAAGCTACCAACCAATCATCTTCAACAGTGCCTCCAAACAGAGAC 833
Db 848 TCTGGCCTCAACAAAGTCTCCCTTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 907
QY 834 CAGCGTCTCAGTATGATCTTATCAGATTTTGGACGGACAGCAGCCGCTCTTGGAAAT 893
Db 908 CAACGGCCCCAGCAGATTCCTATCAGATCTCGGGCCCGACACAGCAGTCGCTTACGCAAC 967
QY 894 CCAGGAGTGGGAGATACAGTATGGCAGTTCCTTACTTGGAGCTTCTGTGGACAGCTCC 953
Db 968 CTTGGAAGCGGAGATCCAGCTGTGCAATCTCTCTGGAGCTGCTCTCCGACAGGCC 1027
QY 954 AACTCCAACTGCATCCTCGGAGGGCACAATGGGAGTTCAAGATGACAGACCCGTAT 1013
Db 1028 AACGCCAGCTGTATCACCTGGGAGGGACCAACGGGGAGTTCAAAATGACGGACCCCGAT 1087
QY 1014 GAAGTGGCTCGGCTGGGGAGAGAGAAAGCAAACTTACATGAATGACAAACTC 1073
Db 1088 GAGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGATATGACAAAGCTG 1147
QY 1074 AGCGTGCACTTGGCTACTATGACAAAAATATTATGACTAAAGTTCTATGTAACGC 1133
Db 1148 AGCGGGCCCTCCGTTATTACTATGATAAAAAATATTGACAAAGTGCAAGGCAAAAGA 1207
QY 1134 TATGCTCAAAATTTGATTTCCAGGATCCTCAGGCTCCAGCCCTCAGCCTCCAGAA 1193
Db 1208 TATGTTTCAAAATTTGATTTCCAGGCAATGCCAGGCTCTGCGAGCCATCCGACCGAG 1267
QY 1194 TCATCCATGTACAAATACCATCAGACCTCCCTACATGAGTTCTTACATGACACCCC 1253
Db 1268 TCGTCCATGTACAAATACCATCTGACATCTCTTACATGCTCTTACCATGCTCCACG 1327
QY 1254 CAGAAGATGAATTTGTAGTCTCCCATCCCGCTGTTGGCCGTAACCTCATCAGCTTT 1313
Db 1328 CAGAAGTGAATTTGTCTCCCTCCCATCTGACATCTCTTACATGCTCTTACCATGCTCC 1387
QY 1314 TTTGCTGCCCTTAATCCATCTGGAATTCACCAACTGGAGGCACTTACCCCAATACCA-- 1371
Db 1388 TTTGAGCGGCAATCAATATCTGACCTCTCCCGAGGGGAACTTACCCCAATACCA-- 1447
QY 1372 -----GGCTGCCAGTCTCATATGCTTCCCATCTTGGACCTACTACTA 1417
Db 1448 GTCCCCCGCCATCTTAACACCCACGTCGCTTACACTTTAGGAGCTACTACTA 1500

RESULT 15
ID ADA02692 standard; cDNA; 1359 BP.
XX
AC ADA02692;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human FLII carcinoma associated coding sequence, SEQ ID NO:1210.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ss.
XX
OS Homo sapiens.
XX
FN WO2003057146-A2.
XX
PD 17-JUL-2003.

```

```

XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
WP1; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
PS useful for preparing compositions for treating carcinomas.
XX
Claim 1; SEQ ID NO 1210; 245pp; English.
XX
The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1359 BP; 359 A; 422 C; 333 G; 245 T; 0 U; 0 Other;
Query Match 38.9%; Score 562.2; DB 8; Length 1359;
Best Local Similarity 66.4%; Pred. No. 16-173;
Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;
QY 63 ATGGCAAGCAGTATTAAAGGAGCATTTATCAGTGGTGAAGTGAAGACAGTCTTTGTTGAG 122
Db 1 ATGGACGGGACTATTAAAGGAGGCTCTGCTGGTGTGAGCGAGCAGTCCCTCTTTTGAC 60
QY 123 TGTGCTTACGGATCG---CCCCACCTTGCAAGACAGAAATGACAGCTCTCTTCCAGT 179
Db 61 TCAGCGTACGGAGCGGCGAGCCCATCTCCCAAGGCGCAGATGACTGCTCGGGAGTCT 120
QY 180 GAATATGGGCACAAACATCAAGATGAGCCGCGCTTCCCCAGCAGACATGGTTATCACAG 239
Db 121 GACTACGGGCGAGCCGCCACAGATCAACCCCTCCACACAGCAGAGTGGATCAATCAG 180
QY 240 CCCCCGGCAGAGTTACCATTAAGATGAGTGTACCCAAACAGAGTAAATGGGTCAAG 299
Db 181 CC-----AGTGAGGGTCAACGTCAAGCGGAGTATGACCAATGAATGATCCAGG 231
QY 300 AATTACCTGATGACTGACGCGTGGCAAAAGAGGAGGAAAATGTTAGCAGTTCAGACAAT 359
Db 232 GAGTCTCCGGTGGACTGACGCGTTAGCAAAATGCAGCAAGCTGTGGCGGAGCGAGTCC 291
QY 360 GTTGGGATGAAGTATGGAAGCTACATGGAGAGAGAGCAT---ATTCCGCTCCAAATATG 416
Db 292 AACCCCATGAAGTACAAACAGCTATATGGACGAGAAAGATGGCCCCCTCTCTCCCAATG 351
QY 417 ACAACCAATGAACGAAGAGTTATTGTGCCAGCAGATCTTACGTTATGGAGCACAGACCAT 476
Db 352 ACCACCAAGAGAGAGAGTATCTGTCCTCCGCGAGACCCACACTGTGGACAGAGGAGCAT 411

```

QY 477 GTACGGCAGTGGCTGGAGTGAAGAGTATGCTTCCAGACGTGGACATCTTG 536
 Db 412 GTGAGGCAATGGCTGGAGTGGCCATAAAGAGTACAGCTTGATGGAGATCGACACATCC 471
 QY 537 TTGTTCCAGAACTTGATGGGAAAGAGTTGTAAATATGACCAAAAGATGACTTCCAGAGA 596
 Db 472 TTTTTCAGAACTGATGGGAAAGAGTGTAAATATGACCAAAAGAGTACTTCTCCGC 531
 QY 597 CTCACGGCAGCTATAACGACAGATATCTCTGTACACCTTACACTACCTCAGAGAGAGA 656
 Db 532 GCCACACCTCTACAACACGGAAGTGTGTGTACACCTTACCTCAGGGAAG- 590
 QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAGCAACGCAAGAAATAACA 716
 Db 591 -----TTCACGTGCTGGCTATAATACACCTCCACACCGACCAATCCTCACGATTGAGT 645
 QY 717 ACAAGCCAGATTTTACCTTATGAGCAAGCAGGAGATCAGGCTGACGAGTACAGCCAT 776
 Db 646 GTCAAAGAGAACCTTCTTATGACTCAGTCAAGAGAGGCTTGGGGCAATACATGAAT 705
 QY 777 CC---CACTCAGTCAAAAGCTACCCAACTATCTTCAACAGTGCCTCAAAACAGAGAC 833
 Db 706 TCTGSCCTCAACAAAGTCTCCCTTGGAGGGGCAAAAGATCAGTAAGATACAGAG 765
 QY 834 CAGGCTCTCAGTTAGATCTTATCAGATTCTTGGACGGACACAGCGCTTGGCAAT 893
 Db 766 CAACGGCCCCAGCAGATCCGTATCAGATCTTGGGCCCCGACACGACGATGCGCTAGCCAC 825
 QY 894 CCAGGAGTGGCGAGATACAGTATGSCAGTTCTTCTGAGAGCTTCTGTGGACAGCTCC 953
 Db 826 CCTGGAACGGCAGATCCAGCTGTGGCAATCTCTCTGAGAGTCTCTCCGACAGCGCC 885
 QY 954 AACTCCAACTGCATCACCTGGAGGGGCAAAATGGGAGTTCAAGATGACAGACCTGAT 1013
 Db 886 AACGCCAGCTATCACCTGGAGGGGACCAACGGGGAGTTCAAAATGACGGACCCCGAT 945
 QY 1014 GAAGTGGCTCGGCTGGGAGAGAGGAAAGCAAACTTACATGAATGACAAATC 1073
 Db 946 GAGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGAATTCAGACAGCTG 1005
 QY 1074 AGCGGTGCACTTCGCTACTATGACAAATAATATGACTAAAGTTTCAATGTPAAACGC 1133
 Db 1006 AGCGGGCCCTCCGTTATTACTATGATAAAACATTTATGACCAAGTGCACGGCAAAAGA 1065
 QY 1134 TATGCTACAAATTTGATTTCCAGGATCGCTCAGGCGCTCCAGCGCTCAGCTCCAGAA 1193
 Db 1086 TATGCTTACAAATTTGATTTCCAGGATTCAGGCGCTTCAGGCGCTTCAGGCGCAATCCGACCGAG 1125
 QY 1194 TCATCCATGTACAAATACCATCAGACCTCCCTTACATGAGTTCTTACCATGACACCCC 1253
 Db 1126 TCGTCCATGTACAGTACCTTCTGACATCTCTACATGCTTCTTACCATGCCACCGAG 1185
 QY 1254 CAGAGATGAATTTGTAGCTCCCATCCCTGCTTGGCCGTAACTCATCCAGCTTT 1313
 Db 1186 CAGAGGTGAATTTGTCTCCCTCCCATCCATCTCCATGCTGCTTCTTCTCCAGCTTC 1245
 QY 1314 TTTGCTGCCCTTATCCATATCGAAATTCACCAACTGGAGGATCTACCCCAATACCA-- 1371
 Db 1246 TTTGGAGCGCATCACATATCTGGACCTCCCGACGGGGGAATCTACCCCAACCCCAAC 1305
 QY 1372 -----GGCTGCCAGCTGCTCATATGCTTCCCATCTTGGCACCTACTACTA 1417
 Db 1306 GTCCCCGCCCATCTTAACACCCAGTGCCTTTCACACTTAGGCAGCTACTACTA 1358

Search completed: May 27, 2004, 16:39:39
 Job time : 687 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2004, 16:28:07 ; Search time 143 Seconds
(without alignments)
5615.482 Million cell updates/sec

Title: US-09-902-772-1
Perfect score: 1447
Sequence: 1 gaattccggaacgaataat.....gaaagaagcgccaagaaaa 1447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6C_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/6D_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1447	100.0	1447	3	US-08-878-177-1
2	1356	93.7	1528	3	US-08-878-177-3
3	563.8	39.0	2938	2	US-08-343-443B-3
4	180	12.4	1752	3	US-09-360-779-1
5	180	12.4	1752	4	US-09-435-335-1
6	148.4	10.3	2532	4	US-09-620-312D-869
7	124.6	8.6	1604	1	US-08-306-691B-43
8	124.6	8.6	1604	5	PCT-US93-06251-9
9	122.2	8.4	2268	3	US-09-344-579-1
10	117.4	8.1	2667	2	US-08-469-412A-1
11	117.4	8.1	2667	2	US-09-021-715-1
12	115.2	8.0	328	2	US-08-343-443B-5
13	100.6	7.0	1933	4	US-09-920-759-3
14	100.6	7.0	1976	4	US-09-920-759-10
15	96	6.6	2064	3	US-08-875-944B-1
16	96	6.6	2064	3	US-09-116-049-3
17	96	6.6	2064	4	US-03-602-868A-1
18	96	6.6	2064	4	US-09-884-363-3
19	95.2	6.6	2410	2	US-08-780-835B-1
20	95.2	6.6	2410	3	US-09-303-268-1
21	95.2	6.6	2410	3	US-09-116-049-1
22	95.2	6.6	2410	4	US-09-884-363-1
23	95.2	6.6	2410	4	US-09-328-925-13
24	93.4	6.5	665	2	US-09-920-759-11
25	88.6	6.1	2266	2	US-09-213-767-1
26	87.4	6.0	2544	2	US-08-469-412A-6
27	87.4	6.0	2544	3	US-09-021-715-6

RESULT 1

US-08-878-177-1
; Sequence 1, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; CURRENT FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA
US-08-878-177-1

Query Match 100.0%; Score 1447; DB 3; Length 1447;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAATTCGCGGAACGAATAATTATTATTAGCAATATTATTAGCGATCAATATCTTGATCACA	60
DB	1	GAATTCGCGGAACGAATAATTATTATTAGCAATATTATTAGCGATCAATATCTTGATCACA	60
QY	61	TTATGGCAAGCACTATTAAAGGAAGCATTTATCATGTGTGAGTGAAGACCGAGTCTTGTTG	120
DB	61	TTATGGCAAGCACTATTAAAGGAAGCATTTATCATGTGTGAGTGAAGACCGAGTCTTGTTG	120
QY	121	AGTGTGCTTACGATCGCCACCTTGCAAGACAGAATGACAGCTCTCTTCAGTG	180
DB	121	AGTGTGCTTACGATCGCCACCTTGCAAGACAGAATGACAGCTCTCTTCAGTG	180
QY	181	AATATGGCAACATCAAGATGACCGCGGCTTCCCGACGAGCTGTGTTATCACAGC	240
DB	181	AATATGGCAACATCAAGATGACCGCGGCTTCCCGACGAGCTGTGTTATCACAGC	240
QY	241	CCCCGCCAGAGTTACCATTAAGATGGAGTGAACCCAAACAGGTTAATGGGTCAAGCA	300
DB	241	CCCCGCCAGAGTTACCATTAAGATGGAGTGAACCCAAACAGGTTAATGGGTCAAGCA	300
QY	301	ATTACCTGATGACTGCAGCGTGGCAAAAGGAGGAAATGTTAGCAGTTGAGCAATG	360
DB	301	ATTACCTGATGACTGCAGCGTGGCAAAAGGAGGAAATGTTAGCAGTTGAGCAATG	360
QY	361	TTGGGATGAACCTATGGAGCTACATGGAGAGCATATTCGGCTCCAAATATGACAA	420
DB	361	TTGGGATGAACCTATGGAGCTACATGGAGAGCATATTCGGCTCCAAATATGACAA	420

ALIGNMENTS

28	79.2	5.5	15788	4	US-09-920-759-13	Sequence 13, Appl
29	67.6	4.7	65042	4	US-09-784-316-3	Sequence 3, Appl
30	59.8	4.1	852	3	US-09-020-956-44	Sequence 44, Appl
31	59.8	4.1	852	3	US-09-030-607-44	Sequence 44, Appl
32	59.8	4.1	852	4	US-09-439-313-44	Sequence 44, Appl
33	59.8	4.1	852	4	US-09-352-616A-44	Sequence 44, Appl
34	59.8	4.1	852	4	US-09-232-149A-44	Sequence 44, Appl
35	59.8	4.1	852	4	US-09-159-812-44	Sequence 44, Appl
36	59.8	4.1	852	4	US-09-636-215-44	Sequence 44, Appl
37	59.8	4.1	852	4	US-09-685-166A-44	Sequence 44, Appl
38	59.8	4.1	852	4	US-09-115-453-44	Sequence 44, Appl
39	59.8	4.1	852	4	US-09-688-489-44	Sequence 44, Appl
40	58.2	4.0	5427	3	US-09-009-913-2	Sequence 3, Appl
41	58.2	4.0	5510	3	US-09-009-913-3	Sequence 3, Appl
42	58.2	4.0	5667	3	US-09-009-913-4	Sequence 4, Appl
43	55.2	3.8	2375	1	US-08-368-281-1	Sequence 1, Appl
44	55.2	3.8	3240	1	US-08-368-281-3	Sequence 3, Appl
45	53.6	3.7	1907	4	US-09-300-958A-27	Sequence 27, Appl


```
QY 421 CCAATGAACGAGAGTTATTGTCGACGAGATCTCTAGTTATGGAGCAGACCATGTAC 480
Db 421 CCAATGAACGAGAGTTATTGTCGACGAGATCTCTAGTTATGGAGCAGACCATGTAC 480
QY 481 GGCAGTGGCTGGAGTGGGACGAGTGAAGAGTATGTCCTCCAGAGCTGAGATCTTTGTTGT 540
Db 481 GGCAGTGGCTGGAGTGGGACGAGTGAAGAGTATGTCCTCCAGAGCTGAGATCTTTGTTGT 540
QY 541 TCCAGAACATTTGATGGGAAAGAGTTGTAAATGACCAAGATGACTTCCAGAGACTCA 600
Db 541 TCCAGAACATTTGATGGGAAAGAGTTGTAAATGACCAAGATGACTTCCAGAGACTCA 600
QY 601 GCGCAGCTATAACGACATATCTCTCTGTCCACACCTACACTACTCTCAGAGAGAGGAG 660
Db 601 GCGCAGCTATAACGACATATCTCTCTGTCCACACCTACACTACTCTCAGAGAGAGGAG 660
QY 661 CCACCTTTTATTTTCCAAATACATCACTGTTTACCAGAAAGCAACGCAAGAAATACACAA 720
Db 661 CCACCTTTTATTTTCCAAATACATCACTGTTTACCAGAAAGCAACGCAAGAAATACACAA 720
QY 721 GGCAGATTTTACCTTATGAGCAACGAGGAGATCAGCGTGCACGAGTCAAGCATCCCA 780
Db 721 GGCAGATTTTACCTTATGAGCAACGAGGAGATCAGCGTGCACGAGTCAAGCATCCCA 780
QY 781 CTCAGTCAAAAGTACCCACCATCATCTTCAACAGTGCACCAAAACAGAGACGAGGTC 840
Db 781 CTCAGTCAAAAGTACCCACCATCATCTTCAACAGTGCACCAAAACAGAGACGAGGTC 840
QY 841 CTCAGTTAGATCTTATCAGATCTTGGACCGACGAGCGCTTGTGCAAAATCCAGGGA 900
Db 841 CTCAGTTAGATCTTATCAGATCTTGGACCGACGAGCGCTTGTGCAAAATCCAGGGA 900
QY 901 GTGGCAGATACAGCTATGGCAGTCTCTACTGGAGTCTCTGTGGACAGCTCCAACTCCA 960
Db 901 GTGGCAGATACAGCTATGGCAGTCTCTACTGGAGTCTCTGTGGACAGCTCCAACTCCA 960
QY 961 ACTGCATCAGCTGGAGGACAAATGGGAGTTCAAGATCAGACGCTCTGATGAGTGG 1020
Db 961 ACTGCATCAGCTGGAGGACAAATGGGAGTTCAAGATCAGACGCTCTGATGAGTGG 1020
QY 1021 CTCGGCGTTGGGAGAGAGGAAAGCAACCTAATGAATGATGACAAACTCAGCCGCTG 1080
Db 1021 CTCGGCGTTGGGAGAGAGGAAAGCAACCTAATGAATGATGACAAACTCAGCCGCTG 1080
QY 1081 CACTTCGCTACTACTACACAAAATATTATGACTAAAGTTCTATGTTAAACGCTATGCT 1140
Db 1081 CACTTCGCTACTACTACACAAAATATTATGACTAAAGTTCTATGTTAAACGCTATGCT 1140
QY 1141 ACAAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCCCTCCAGAAATCATCCA 1200
Db 1141 ACAAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCCCTCCAGAAATCATCCA 1200
QY 1201 TGTTCAAAATACCATCAGACCTCCGCTACATGAGTTCCTACCATGCAACCCCGAGAGA 1260
Db 1201 TGTTCAAAATACCATCAGACCTCCGCTACATGAGTTCCTACCATGCAACCCCGAGAGA 1260
QY 1261 TGAACCTTTGTAGTCCCATCCCTGCTTTGCGCGCTAACTCATCCAGCTTTTGTGCTG 1320
Db 1261 TGAACCTTTGTAGTCCCATCCCTGCTTTGCGCGCTAACTCATCCAGCTTTTGTGCTG 1320
QY 1321 CCCCTAATCCATCTGGAATTTCAACACTGAGGCACTTACCCCAATACAGGCTGCCAG 1380
Db 1321 CCCCTAATCCATCTGGAATTTCAACACTGAGGCACTTACCCCAATACAGGCTGCCAG 1380
QY 1381 CTGCTCATATGCTTCCCATCTTGGCAGCTACTACTAAGTGGGAAAGAAAGAGCGCC 1440
Db 1381 CTGCTCATATGCTTCCCATCTTGGCAGCTACTACTAAGTGGGAAAGAAAGAGCGCC 1440
QY 1441 AAGAAAA 1447
Db 1441 AAGAAAA 1447
```

RESULT 2

```
US-08-878-177-3
; Sequence 3, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Seliyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; TITLE OF INVENTION: the Proteins
; FILE REFERENCE: chugai seliyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: DNA
; ORGANISM: c-erg gene, chicken DNA
US-08-878-177-3

Query Match 93.7%; Score 1356; DB 3; Length 1528;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1447; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

QY 1 GAATTCGCGAAGCAATTAATTTATAGCAATATTAGCGATCAATATCTTGATCACA 60
Db 1 GAATTCGCGAAGCAATTAATTTATAGCAATATTAGCGATCAATATCTTGATCACA 60
QY 61 TTATGGCAAGCACTATTAAAGGAAGCATTTATCAGTGGTGAAGCAAGCAGTCTTTGTTG 120
Db 61 TTATGGCAAGCACTATTAAAGGAAGCATTTATCAGTGGTGAAGCAAGCAGTCTTTGTTG 120
QY 121 AGTGTGCTACGGATCGCCCACTTGAAGAGCAAGATGCAAGAGAGAAATGACAGCTTCCAGTG 180
Db 121 AGTGTGCTACGGATCGCCCACTTGAAGAGCAAGATGCAAGAGAGAAATGACAGCTTCCAGTG 180
QY 181 AATATGGGCAAAACATCAAAAGATGAGCCGCGCTTCCCGAGCAGGACTGGTTATCACAGC 240
Db 181 AATATGGGCAAAACATCAAAAGATGAGCCGCGCTTCCCGAGCAGGACTGGTTATCACAGC 240
QY 241 CCCCGCCAGAGTTACCAATTAAGATGGAGTGAACCCAAAACAGGTTAATGGGTCAAGGA 300
Db 241 CCCCGCCAGAGTTACCAATTAAGATGGAGTGAACCCAAAACAGGTTAATGGGTCAAGGA 300
QY 301 ATTCACTGTATGATGACATGCAAGAGTGAAGAGGAGAAATAGTTAGCAAGTTCCAGACAATG 360
Db 301 ATTCACTGTATGATGACATGCAAGAGTGAAGAGGAGAAATAGTTAGCAAGTTCCAGACAATG 360
QY 361 TTGGGATGAATCTATGGAAGCTACATGGAAGAGAGACATATTCGCGCTCCAAATATGACAA 420
Db 361 TTGGGATGAATCTATGGAAGCTACATGGAAGAGAGACATATTCGCGCTCCAAATATGACAA 420
QY 421 CCAATGACGAGAGTTATTTGTCGACGAGATCTTACGTTATGGAGCAGACACCATGTAC 480
Db 421 CCAATGACGAGAGTTATTTGTCGACGAGATCTTACGTTATGGAGCAGACACCATGTAC 480
QY 481 GGCAGTGGCTGGAGTGGGAGTGAAGAGTATGTCCTTCCAGACCTGAGATCTTTGTTGT 540
Db 481 GGCAGTGGCTGGAGTGGGAGTGAAGAGTATGTCCTTCCAGACCTGAGATCTTTGTTGT 540
QY 541 TCCAGAACATTTGATGGGAAAGAGTTGTAAATGACCAAGATGACTTCCAGAGACTCA 600
Db 541 TCCAGAACATTTGATGGGAAAGAGTTGTAAATGACCAAGATGACTTCCAGAGACTCA 600
QY 601 GCGCAGCTATAACGACATATCTCTCTGTCCACACCTACACTACTCTCAGAGAGAGAG 654
Db 601 GCGCAGCTATAACGACATATCTCTCTGTCCACACCTACACTACTCTCAGAGAGAGAG 660
QY 655 ----- 654
Db 661 TTCCACATTTGACTTTCAGATGATGTGATAAGCGCTTACAAAACTCTCCACGGTTAATGC 720
QY 655 -----GAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAGAG 699
```

Db 721 ATGCTAGAAAACACAGGAGGAGCCATTTTATTTTCCAAATACATCAGTTTACCAGAAG 780
Qy 700 CAACGCAAGATACACAAAGCCAGATTTACCTTATGCAAGCGAGGAGATCAGCGT 759
Db 781 CAACGCAAGATACACAAAGCCAGATTTACCTTATGCAAGCGAGGAGATCAGCGT 840
Qy 760 GGACGAGTCACAGCCATCCCACTCAGTCAAAAGTACCCAAACCATCATCTTCAACAGTGC 819
Db 841 GGACGAGTCACAGCCATCCCACTCAGTCAAAAGTACCCAAACCATCATCTTCAACAGTGC 900
Qy 820 CCAAAACAGAGACACAGCGTCTCAGTTAGATCTTATCAGATCTTGGACCGACGCA 879
Db 901 CCAAAACAGAGACACAGCGTCTCAGTTAGATCTTATCAGATCTTGGACCGACGCA 960
Qy 880 GCGGCTCTTCAAACTCCAGGAGTGGGAGATACAGCTATGCGAGTCTTCTACTGGAGTTC 939
Db 961 GCGGCTCTTCAAACTCCAGGAGTGGGAGATACAGCTATGCGAGTCTTCTACTGGAGTTC 1020
Qy 940 TGTGGAGAGCTCCAACTCCAACTGATCCTCTGGAGGGGACAAATGGGAGTTCAGA 999
Db 1021 TGTGGAGAGCTCCAACTCCAACTGATCCTCTGGAGGGGACAAATGGGAGTTCAGA 1080
Qy 1000 TGACAGACCTGATGAAGTGGCTCGCGTTGGGAGAGAGAAAGCAAACTTACATGA 1059
Db 1081 TGACAGACCTGATGAAGTGGCTCGCGTTGGGAGAGAGAAAGCAAACTTACATGA 1140
Qy 1060 ACTATGACAACTCAGCGGTGACCTGCTACTATGACAAATAATTTATGACTAAG 1119
Db 1141 ACTATGACAACTCAGCGGTGACCTGCTACTATGACAAATAATTTATGACTAAG 1200
Qy 1120 TTCATGGTAAAGCTATGCTTCAAAATTTGATTTCCAGGAGTCTCAGGCGCTCCAGC 1179
Db 1201 TTCATGGTAAAGCTATGCTTCAAAATTTGATTTCCAGGAGTCTCAGGCGCTCCAGC 1260
Qy 1180 CTCACCTCCAGATCATCCATGTACAAATACCCATCAGACCTCCCTTACATGATCTCT 1239
Db 1261 CTCACCTCCAGATCATCCATGTACAAATACCCATCAGACCTCCCTTACATGATCTCT 1320
Qy 1240 ACCATGACACCCGAGAGATGACTTTGTAGCTCCCATCCCTGCTTTGCCGTA 1299
Db 1321 ACCATGACACCCGAGAGATGACTTTGTAGCTCCCATCCCTGCTTTGCCGTA 1380
Qy 1300 CTTATCCAGCTTTTGTGCTCCCTTATCCATCTGGAATTCACCAACTGAGGCACTCT 1359
Db 1381 CTTATCCAGCTTTTGTGCTCCCTTATCCATCTGGAATTCACCAACTGAGGCACTCT 1440
Qy 1360 ACCCAATACAGGCTGCGAGTCTCATATGCTTCCCATCTTGGACCTTACTACTAAG 1419
Db 1441 ACCCAATACAGGCTGCGAGTCTCATATGCTTCCCATCTTGGACCTTACTACTAAG 1500
Qy 1420 TGGGGAAGAAAGAAAGCGCCAGAAA 1447
Db 1501 TGGGGAAGAAAGAAAGCGCCAGAAA 1528

RESULT 3

US-08-343-443B-3
; Sequence 3, Application US/08343443B
; Patent No. 598734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Melot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploegastel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF
; CANCEROUS
; TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID

; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989.6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2938 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 143..1498
; US-08-343-443B-3

Query Match 39.0%; Score 563.8; DB 2; Length 2938;
Best Local Similarity 56.5%; Pred. No. 1.2e-183;
Matches 913; Conservative 0; Mismatches 427; Indels 33; Gaps 6;
Qy 63 ATGCGACGACTATTAGGAGGAGTATCAGTGGTGAAGACGAGTCTTGTGTTGAG 122
Db 143 ATGCGAGGAGTATTAGGAGGCTCTGTGGTGGTGAAGACGAGTCTTGTGTTGAG 202
Qy 123 TGTGCTTACGAGTCG---CCCCACCTTGCAGGAGAGAAATGACAGCTCTTTCAGT 179
Db 203 TCAGGTACGAGGCGGCGAGCCCATCTCCCAAGGCGAGCATGCTGCTCGGGAGTCT 262
Qy 180 GAATATGGCAACATCAAGATGAGCGCGGCTTCCCGAGGAGTGGTTATCAGAG 239
Db 263 GACTACGGGAGCGCCCAAGATCAACCCCTCCACACGAGGAGTGGATCAATCAG 322
Qy 240 CCCCAGCCAGAGTTACCATTAAGATGGAGTGAACCCAAACAGGTTAATGGTCAAG 299
Db 323 CC-----AGTAGGCTCAACGTCAGCGGGAGTATGACCACATGAATGATCCAG 373
Qy 300 AATTCACCTGATGCTCAGCGTGGGCAAGAGGAGGAAATGTTAGTCAGTTCAGACA 359
Db 374 GAGTCTCCGTTGAGTCTGACGCTTAGCAATGAGCAAGCTGTGGCGGAGGAGTCC 433
Qy 360 GTTGGGATGAATGAGTGAAGCTACATGGAAGAGAGCAT---ATTCCGCTCCAAATATG 416
Db 434 AACCCATGAATACACACATATATGACGAGAGAGATGCGCCCTCTCTCCCAACATG 493
Qy 417 ACAACCAATGAACGAGAGTATTGTCCAGAGAGATCTTCTAGTATGGAGCAGACCAT 476
Db 494 ACCACCAAGAGAGAGTATGCTCCCGAGACCCCACTGTGTGGACAGGAGCAT 553

477 GTACGGAGTGGCTGGAGTGGCGAGTGAAGAGTATGTTCTTCCAGAGCTGGACATCTTG 536
Db GTGAGGCAATGGCTGGAGTGGCGCCATAAAGCAGTATAGCTTGTATGGAGATCGACACATCC 613
QY TTGTTCCAGACATGATGGAAAGAGTGTGTTAAATGACCAAGATGATCTTCCAGAG 596
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 673
QY CTCACGGCGAGTATAACGCGAGATATCTCTGTCCACACCTACACTACCTCAGAGAGAG 656
Db GCCACACCTCTACAAACCGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 732
QY GGAGCCACTTTATTTTCCAAATACATCATGTTTACCCGAGAGCAACGCAAGATTAACA 716
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 787
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 776
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 847
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 833
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 907
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 893
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 967
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 953
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1027
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1013
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1087
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1073
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1147
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1133
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1207
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1193
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1267
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1253
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1327
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1313
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1387
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1371
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1447
QY TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1417
Db TTTTCCAGAACATGATGGAAAGAGTGTGTTAAATGACCAAGAGGAGTCTTCTCCCGC 1500

RESULT 4

US-09-360-779-1

; Sequence 1, Application US/09360779

; Patent No. 628216

; GENERAL INFORMATION:

; APPLICANT: Deneris, Evan S.

; APPLICANT: Fyodorov, Dmitry V.

APPLICANT: Hendricks, Timothy J.
TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
FILE REFERENCE: Useful in the Treatment of Neurological Diseases
CURRENT APPLICATION NUMBER: US/09/360,779
CURRENT FILING DATE: 1999-07-26
EARLIER APPLICATION NUMBER: 60/094,264
EARLIER FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1752
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
NAME/KEY: CDS
LOCATION: (112)..(1131)
US-09-360-779-1

Query Match 12.4%; Score 180; DB 3; Length 1752;

Best Local Similarity 74.0%; Pred. No. 3.2e-51;

Matches 228; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 879 AGCGTCTTTGCAAAATCCAGGAGTGGCGAGATACAGATCCAGTTCCTACTGGAGCTT 938
Db 529 AGCGTCTTCGCTACAGAAAGGCGAGATCCAGTTCCTACTGGAGCTG 588
QY 939 CTGTCGGACAGCTCCAACTCCAACTGCACTCGGAGGGCACAATATGGGAGTTCAG 998
Db 589 CTGTCGGACAGCTCCGAGAACCGCGCTGCATCGCTGGAGGGGGCCACGCGAGTTCAG 648
QY 999 ATGACAGACCTTCATGAAAGTGGCTCGGCTGGGAGAGAGGAAAGCAAACTTACATG 1058
Db 649 CTCACCGACCCCGACGAGGTGGCGACGCTGGGCGAGCGCAAGAGCCCAATATG 708
QY 1059 AACTATGACAAATCAGCGCTGCACTTCGCTACTACTGACAAAATATGACTAAA 1118
Db 709 AACTATGACAAATCAGCGCTGCACTTCGCTACTACTGACAAAATATGACTAAA 768
QY 1119 GTTCATGTAAGAGCTATGCTTACAAATTTGATTTCCAGGAATCGCTCAGGCTCCAG 1178
Db 769 GTGACGGCAAGCGCTACGCTTACGCTTTCAGCTTCCAGGGCTGCGACAGGCTTCCAG 828
QY 1179 CTTACCC 1186
Db 829 CCACCACC 836

RESULT 5

US-09-435-335-1

; Sequence 1, Application US/09435335

; Patent No. 6384204

; GENERAL INFORMATION:

; APPLICANT: Deneris, Evan S.

; APPLICANT: Fyodorov, Dmitry V.

; APPLICANT: Hendricks, Timothy J.

; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds

; FILE REFERENCE: Useful in the Treatment of Neurological Diseases

; CURRENT APPLICATION NUMBER: US/09/435,335

; CURRENT FILING DATE: 1999-11-05

; EARLIER APPLICATION NUMBER: 09/360,779

; EARLIER FILING DATE: 1999-07-26

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1752

; TYPE: DNA

; ORGANISM: Rattus norvegicus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (112)..(1131)

; US-09-435-335-1

Query Match 12.4%; Score 180; DB 4; Length 1752;
Best Local Similarity 74.0%; Pred. No. 3.28-51;
Matches 228; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 879 AGCGTCTTGGCAATCCAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGGAGCTT 938
Db 529 AGCCCTGGGTACAGAAAGGAGCGGGCAGATCCAGTTGTGGCAGTTTCTACTGGAGCTG 598
QY 939 CTGTCGGACAGCTCCAACTCCAACTGCATCCTGGGAGGCGCAAAATGGGGAGTTCAAG 998
Db 599 CTGGCAGACCGCGGAACCGCGGTGCATCGCGTGGGAGGCGGCCACCGCGAGTTCAAG 648
QY 999 ATGACAGACCTGTATGAAGTGGCTGGCGTTGGGGAGAGAGAAAAGCAAACTTAACATG 1058
Db 649 CTCACGACCCGACGAGTGGCGGACGCTGGGCGGAGCGCAGAGAGGCCCAATATG 708
QY 1059 AACTATGACAACTCAGCGGTGCACTTGGCTACTACTATGATGACAAAATATTATGACTAAA 1118
Db 709 AACTACGACAAAGTAAAGTCGAGCACTGGCTACTACTACGACAAAACATCATGAGCAAG 768
QY 1119 GTTCATGTAAAGCTATGCTTACAAATTTGATTTCACCGGAATCGCTCAGGCCCTCCAG 1178
Db 769 GTGCACGCAAGCGCTACGCTACCGCTTTCACCTTCAGGCGCTGGCAGAGGCTTGCCAG 828
QY 1179 CCTCACCC 1186
Db 829 CCACCACC 836

RESULT 6
US-09-620-312D-869
; Sequence 869, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; PRIOR FILING DATE: 2000-07-19
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: DT_FL_genes Version 1.0
; SEQ ID NO 869
; LENGTH: 2532
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2532)
US-09-620-312D-869

Query Match 10.3%; Score 148.4; DB 4; Length 2532;

Best Local Similarity 73.5%; Pred. No. 3.4e-40;
Matches 216; Conservative 0; Mismatches 76; Indels 2; Gaps 2;
QY 891 AATCCAGGAGTGGGAGATACAGCTATGGGAGTTCCTACTGGAGCTTCTGTCGACAGC 950
Db 2091 AATGATGAGGCGGAGAGATCCAGCTGTGGCAATTTCTGCTGGAGCTGCTGGTGAACCGC 2150
QY 951 TCCAACTCCAACTGCATCACCCTGGGAGGCGCAAAATGGGAGTTCAGAGATGACAGACCTT 1010
Db 2151 GCGACGCGG-CTGANTCGGTGGGAGGCGGTCCAGGCGAGTTCAAGCTCAGGACCGG 2209
QY 1011 GATGAAGTGGCTGGCGTTGGGGAGAGAGAAAAGCAAACTTAACATGAACTATGACAAA 1070
Db 2210 GACGAGTGGCGGCGGTGGGCGGAGCGCAAGAGCAAGCCCAACATGAATACGACAAAG 2269
QY 1071 CTCAGCGGTGCACTTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCATGTAAA 1130
Db 2270 CTGAGCGGCGCTTGTAT-CTACTACGACAAACATCATGAGCAAGTGCATGGTAAG 2328
QY 1131 CGCTATGCTTACAAATTTGATTTCACCGGAATCGCTCAGGCCCTCCAGCCTCAC 1184
Db 2329 CGTACGCTACCGCTTCGACTTCAGGCGCTGGGCGAGGCGGCGGCGCTCTAC 2382

RESULT 7
US-08-306-691B-43
; Sequence 43, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Galabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorigna & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: NO. 5734039e
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1604 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-306-691B-43

Query Match 8.6%; Score 124.6; DB 1; Length 1604;
Best Local Similarity 61.6%; Pred. No. 4.1e-32;
Matches 199; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 835 AGCGTCTCAGTTAGATCCTTATCAGATTCTTGGACCGACAGCGCGTCTTGCAGATC 894
Db 1207 ACCGTGCTGACCTCAATAAGGACAAGCGCTGCTATTCTGCTGCTGCCCTAGCTGGGTACA 1266
QY 895 CAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTGGCAGCTCCA 954
Db 1267 CAGGAGTGGACCAATCCAGCTATGCGAGTTCCTTCTGGAACTACTCACTGATAATCCT 1326
QY 955 ACTCCAACTGCATCACCTCTGGAGGCGCAAAATGGGGAGTTTCAAGATGACAGACCCCTGATG 1014
Db 1327 GTCAGTCTTTTATCAGCTGGACAGAGATGGCTGGGAATTCAAACTTTCTGACCCAGATG 1386
QY 1015 AAGTGGCTCGCGTGGGAGAGAGGAAGCAACCTACATGAACTATGACAACTCA 1074
Db 1387 AGGTGGCGAGAGATGGGAAAGAGGAAACCAACCTTAAGATGAATATGAGAACTGA 1446
QY 1075 GCCGTGCACTTCGCTACTACTATGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1134
Db 1447 GCCGTGGCTACGCTACTATTACGACAAATAATTCATCCACAGACAGCGGGAAACGCT 1506
QY 1135 ATGCTACAAATTTGATTTCCAC 1157
Db 1507 ACGTGTACCGCTTTGTGTGAC 1529

RESULT 8
PCT-US93-06251-9
; Sequence 9, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1604 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-9

Query Match 8.6%; Score 124.6; DB 5; Length 1604;
Best Local Similarity 61.6%; Pred. No. 4.1e-32;
Matches 199; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 835 ACCGTCTCAGTTAGATCCTTATCAGATTCTTGGACCGACAGCGCGTCTTGCAGATC 894
Db 1207 ACCGTGCTGACCTCAATAAGGACAAGCGCTGCTATTCTGCTGCTGCCCTAGCTGGGTACA 1266

QY 895 CAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTGGACAGCTCCA 954
Db 1267 CAGGAGTGGACCAATCCAGCTATGCGAGTTCCTTCTGGAACTACTCACTGATAATCCT 1326
QY 955 ACTCCAACTGCATCACCTCTGGGAGGCGCAAAATGGGGAGTTTCAAGATGACAGACCCCTGATG 1014
Db 1327 GTCAGTCTTTTATCAGCTGGACAGAGATGGCTGGGAATTCAAACTTTCTGACCCAGATG 1386
QY 1015 AAGTGGCTCGCGTGGGAGAGAGGAAAGCAACCTTAACATGAATGATGACAACTCA 1074
Db 1387 AGGTGGCGAGAGATGGGAAAGAGGAAACCAACCTTAAGATGAATATGAGAACTGA 1446
QY 1075 GCCGTGCACTTCGCTACTACTATGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1134
Db 1447 GCCGTGGCTACGCTACTATTACGACAAATAATTCATCCACAGACAGCGGGAAACGCT 1506
QY 1135 ATGCTACAAATTTGATTTCCAC 1157
Db 1507 ACGTGTACCGCTTTGTGTGAC 1529

RESULT 9
US-09-344-579-1
; Sequence 1, Application US/09344579
; Patent No. 6054316
; GENERAL INFORMATION:
; APPLICANT: Brenda F. Baker
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF ETS-2 EXPRESSION
; FILE REFERENCE: RTS-0063
; CURRENT APPLICATION NUMBER: US/09/344, 579
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (291)..(1700)
US-09-344-579-1

Query Match 8.4%; Score 122.2; DB 3; Length 2268;
Best Local Similarity 66.5%; Pred. No. 3.6e-31;
Matches 175; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 895 CAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTGGAGAGCTCCA 954
Db 1363 CAGGAGTGGACCTATTTCAGCTGTGGCAGTTTCTCTGGAGCTGCTATCAGACAAATCCT 1422
QY 955 ACTCCAACTGCATCACCTCTGGGAGGCGCAAAATGGGGAGTTTCAAGATGACAGACCCCTGATG 1014
Db 1423 GCCAGTCAATTCATCAGCTGGACTGGAGAGCGGATGGGAGTTTAAAGCTCGCCGCCCGGATG 1482
QY 1015 AAGTGGCTCGCGTGGGAGAGAGGAAAGCAACCTTAACATGAATGATGACAACTCA 1074
Db 1483 AGGTGGCGCGCGTGGGAAAGAGGAAATAATGACCCAGATGAATCTACGAGAGCTGA 1542
QY 1075 GCCGTGCACTTCGCTACTACTATGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1134
Db 1543 GCCGGGCTTACGCTACTATTACGACAAATAATTCATCCACAGACAGCTCGGGGAAACGCT 1602
QY 1135 ATGCTACAAATTTGATTTCCAC 1157
Db 1603 ACGTGTACCGCTTTGTGTGAC 1625

RESULT 10
US-08-469-412A-1
; Sequence 1, Application US/08469412A
; Patent No. 5856125
; GENERAL INFORMATION:

APPLICANT: Mavrothalassitis, George J.
APPLICANT: Blair, Donald G.
APPLICANT: Fisher, Robert J.
APPLICANT: Beal Jr., Gregory J.
APPLICANT: Athanasiou, Meropi A.
APPLICANT: Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS: 16
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,412A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Garrett-Mackowski, Eugenia
REGISTRATION NUMBER: 37,330
REFERENCE/DOCKET NUMBER: 015280-229000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2667 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 123..1769
OTHER INFORMATION: /note= "human ERF (ETS2 Repressor
OTHER INFORMATION: Factor) cDNA"

US-08-469-412A-1
Query Match 8.1%; Score 117.4; DB 2; Length 2667;
Best Local Similarity 63.1%; Pred. No. 1.8e-29;
Matches 181; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 894 CCAGGAGTGGGAGATACAGTATGGCAGTTCCTACTGGAGCTTCTGTGGACAGCTCC 953
DB 186 CTGGCTCAAGGCAGATCCAGCTGTGGCACTTTATCTCTGGAGCTGCTGCGAAGGAGG 245
QY 954 AACTCCAACTGCATCACCCTGGGAGGACAAATGGGAGTTCAAGATGACAGACCTCTGAT 1013
DB 246 TACAGGGCGTTCATTCCTGCGAGGGGACTACGGGGAATTCGTCAACAAGACCTTGAT 305
QY 1014 GAAGTGGCTGGCGTGTGGGAGAGAGAAAGCAAACTAATCACTATGACAAATCTC 1073
DB 306 GAGGTGGCCGGCTGTGGGGGCTTCGCAAGTGCAGAGCCCAAGATGAATTAGCAAGCTG 365
QY 1074 AGCGGTGCATTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCATGGTAAACGC 1133
DB 366 AGCGGGCCCTCGGCTATTACTATACAGCGCAATTCGACAGACCAAGGGAACGG 425
QY 1134 TATGCTCAAAATTTGATTTCCACGGAATCGCTACGCCCTCCAGCC 1180
DB 426 TTCACTCAAGTTCAATTTCAACAAACTGGTGTGTCATTAATACCC 472

RESULT 11
US-09-021-715-1
; Sequence 1, Application US/09021715

Patent No. 6194547
GENERAL INFORMATION:
APPLICANT: Mavrothalassitis, George J.
APPLICANT: Blair, Donald G.
APPLICANT: Fisher, Robert J.
APPLICANT: Beal Jr., Gregory J.
APPLICANT: Athanasiou, Meropi A.
APPLICANT: Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS: 16
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/021,715
FILING DATE: 10-Feb-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Garrett-Mackowski, Eugenia
REGISTRATION NUMBER: 37,330
REFERENCE/DOCKET NUMBER: 015280-229000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2667 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 123..1769
OTHER INFORMATION: /note= "human ERF (ETS2 Repressor
OTHER INFORMATION: Factor) cDNA"

US-09-021-715-1
Query Match 8.1%; Score 117.4; DB 3; Length 2667;
Best Local Similarity 63.1%; Pred. No. 1.8e-29;
Matches 181; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 894 CCAGGAGTGGGAGATACAGTATGGCAGTTCCTACTGGAGCTTCTGTGGACAGCTCC 953
DB 186 CTGGCTCAAGGCAGATCCAGCTGTGGCACTTTATCTCTGGAGCTGCTGCGAAGGAGG 245
QY 954 AACTCCAACTGCATCACCCTGGGAGGACAAATGGGAGTTCAAGATGACAGACCTCTGAT 1013
DB 246 TACAGGGCGTTCATTCCTGCGAGGGGACTACGGGGAATTCGTCAACAAGACCTTGAT 305
QY 1014 GAAGTGGCTGGCGTGTGGGAGAGAGAAAGCAAACTAATCACTATGACAAATCTC 1073
DB 306 GAGGTGGCCGGCTGTGGGGGCTTCGCAAGTGCAGAGCCCAAGATGAATTAGCAAGCTG 365
QY 1074 AGCGGTGCATTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCATGGTAAACGC 1133
DB 366 AGCGGGCCCTCGGCTATTACTATACAGCGCAATTCGACAGACCAAGGGAACGG 425
QY 1134 TATGCTCAAAATTTGATTTCCACGGAATCGCTACGCCCTCCAGCC 1180
DB 426 TTCACTCAAGTTCAATTTCAACAAACTGGTGTGTCATTAATACCC 472

RESULT 12
 US-08-343-443B-5
 ; Sequence 5, Application US/08343443B
 ; Patent No. 5968734
 ; GENERAL INFORMATION:
 ; APPLICANT: Aurias, Alain
 ; APPLICANT: Delattre, Olivier
 ; APPLICANT: Desmaze, Chantal
 ; APPLICANT: Melot, Thomas
 ; APPLICANT: Peter, Martine
 ; APPLICANT: Ploougaastel, Beatrice
 ; APPLICANT: Thomas, Gilles
 ; APPLICANT: Zucman, Jessica
 ; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
 ; CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
 ; TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
 ; TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
 ; TRANSLOCATIONS
 ; NUMBER OF SEQUENCES: 129
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Weiser & Associates
 ; STREET: 230 South Fifteenth Street
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: AEDIT 1.0 DOS text editor
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/343.443B
 ; FILING DATE: 18-NOV-1994
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/FR93/00494
 ; FILING DATE: 19-MAY-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: FR 92/06123
 ; FILING DATE: 20-MAY-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Weiser, Gerard J.
 ; REGISTRATION NUMBER: 19,763
 ; REFERENCE/DOCKET NUMBER: 989.6121P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-875-8383
 ; TELEFAX: 215-875-8394
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 328 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..327
 ; US-08-343-443B-5
 Query Match 8.0%; Score 115.2; DB 2; Length 328;
 Best Local Similarity 67.4%; Pred. No. 2.3e-29;
 Matches 178; Conservative 0; Mismatches 83; Indels 3; Gaps 1;
 QY 733 CTTATGACGACGAGGAGATCAGCGTGGACGAGTCACAGCCATCC---CACTCAGTCAA 789
 DB 65 CTTATGACTCAGTCAGAGAGAGGCTTGGGGCAATAACATGAATCTGGCCCTCAACAAA 124
 QY 790 AAGCTACCCAAACCATCATCTTCAACAGTGGCCCAACAGAGACAGCGTCTCCTCAGTAG 849
 DB 125 GTCTCTCCCTTGGAGGGGCACAAACGATCAGTAAAGATACAGAGCAACGGCCCGACGAG 184
 QY 850 ATTCCTATCAGATTTCTGACCGACACAGCGCTTTCGAAATCCAGGGAGTGGGAGA 909

Db 185 ATCCGTATCAGATCTCTGGGCCGACGACGATCGCTAGCCAAACCTTGAAGGGGCGAGA 244
 QY 910 TACAGCTATGGCAGTTCTTACTTGGAGCTTCTGTGGACAGTCCAACTCCAACTGCATCA 969
 Db 245 TCAGCTGTGGCAATTCCTCTGGAGCTCTCTCCGACAGCGCCAAACGCGAGCTGTATCA 304
 QY 970 CTTGGGAGGGCACAAATATGGGAGT 993
 Db 305 CTTGGGAGGGGACCAACACGGGGAGT 328
 RESULT 13
 US-09-920-759-3
 ; Sequence 3, Application US/09920759
 ; Patent No. 6537811
 ; GENERAL INFORMATION:
 ; APPLICANT: Brenda F. Baker
 ; APPLICANT: Susan M. Freier
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF SAP-1 EXPRESSION
 ; FILE REFERENCE: RTS-0267
 ; CURRENT APPLICATION NUMBER: US/09/920,759
 ; CURRENT FILING DATE: 2001-08-01
 ; NUMBER OF SEQ ID NOS: 91
 ; SEQ ID NO 3
 ; LENGTH: 1933
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (150)...(1445)
 ; US-09-920-759-3
 Query Match 7.0%; Score 100.6; DB 4; Length 1933;
 Best Local Similarity 64.3%; Pred. No. 9.2e-24;
 Matches 151; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
 QY 915 CTATGGCAGTTCTTACTGGAGCTTCTGTGGACAGCTCCAACTCCAACTGCATCACCTGG 974
 Db 168 CTGTGGCAGTTCTTCTTTCAGCTCTCTGAGAAAGCCTCAGAACAGACATGATCTGTGG 227
 QY 975 GAGGGCACAAATGGGGAGTTCAAGATGACAGACCCCTGATGAAGTGGCTCGCGTTGGGA 1034
 Db 228 ACCTCTAATGATGGGCAGTTTAAAGCTTTTTCAGGAGAGAGGTTGGCTCTCTGGGGG 287
 QY 1035 GAGAGAAAAGCAACCTTAACATGAATGAACTATGACAACTCAGCGGTGCATTCGCTACTAC 1094
 Db 288 ATTCCGACAGACAGCCCTAACATGAATTATGACAACTCAGCGGAGCCCTCAGATCTAT 347
 QY 1095 TATGACAAAATATTATGACTAAAGTTTCATGTTAAACGCTATGCTCCCTACAAATTTG 1149
 Db 348 TATGTAAGATATCATCAAAAAGTGAATGTCAGAAAGTTTGTGTACAAGTTG 402
 RESULT 14
 US-09-920-759-10
 ; Sequence 10, Application US/09920759
 ; Patent No. 6537811
 ; GENERAL INFORMATION:
 ; APPLICANT: Brenda F. Baker
 ; APPLICANT: Susan M. Freier
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF SAP-1 EXPRESSION
 ; FILE REFERENCE: RTS-0267
 ; CURRENT APPLICATION NUMBER: US/09/920,759
 ; CURRENT FILING DATE: 2001-08-01
 ; NUMBER OF SEQ ID NOS: 91
 ; SEQ ID NO 10
 ; LENGTH: 1976
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (150)...(1367)
 ; US-09-920-759-10

Query Match 7.0%; Score 100.6; DB 4; Length 1976;
 Best Local Similarity 64.3%; Pred. No. 9.4e-24;
 Matches 151; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 915 CTATGGCAGTTCTCTACTGGAGTTCTCTGCGACAGCTCCAACTGCACTGCACTGG 974
 DB 168 CTGCGCAGTTCTCTCTCTGAGAGCTCTGAGAGCTCTGAGAGCAATGATCTGTTGG 227
 QY 975 GAGGCGCAAAATGGGAGTTCAAGATGACAGACCCCTGATGAGTGGCTGCGCGTTGGGGA 1034
 DB 228 ACCTCTAATGATGGGCGAGTTTAAAGCTTTTGAGGCGAGAGAGTGGCTCTCTCTGGGGG 287
 QY 1035 GAGAGGAAAGCAACCTTAACTGATGACTATGACAACTCAGCCCTGCACTTCGCTACTAC 1094
 DB 288 ATTCCAGAACAGCCCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 347
 QY 1095 TATGCAAAAATATTTATGACTTAAAGTTTCAATGCTGATAAGCTTATGCTTACAAATTTG 1149
 DB 348 TATGTAAGAATATCATCTCAAAAAGTGAATGGTCAGAAGTTTGTGTACAAAGTTTG 402

RESULT 15
 US-08-875-944B-1
 ; Sequence 1, Application US/08875944B
 ; Patent No. 6096542
 ; GENERAL INFORMATION:
 ; APPLICANT: FUJINAGA, Kei
 ; APPLICANT: YOSHIDA, Koichi
 ; APPLICANT: HIGASHINO, Fumihiro
 ; TITLE OF INVENTION: CANCER CONTROL
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: BROWDY AND NEIMARK, P.L.L.C.
 ; STREET: 624 Ninth Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20001
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/875,944B
 ; FILING DATE: 07-AUG-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 07-020173
 ; FILING DATE: 08-FEB-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/00016
 ; FILING DATE: 09-JAN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BROWDY, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: FUJINAGA-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2064 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1386
 ; US-08-875-944B-1

Query Match 6.8%; Score 96; DB 3; Length 2064;
 Best Local Similarity 62.5%; Pred. No. 3.7e-22;
 Matches 150; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 910 TACAGCTATGGCAGTTCTCTACTGGAGTTCTCTGCGACAGCTCCAACTGCAATGCAATCA 969
 DB 956 TGCAGCTGTGGCAATTTCTGTGGTGGCCTTCTGATGACCCAAACAAATGCCATTTCAATTG 1015
 QY 970 CTTGGGAGGCGCAAAATGGGAGTTTCAAGATGACAGACCCCTGATGAGTGGCTGCGGGTT 1029
 DB 1016 CTTGGAGCGGCGCGGGAATGGAGTTTCAAGCTCATTTGAGCTGAGGAGGTGCGCAGGCTCT 1075
 QY 1030 GGGGAGAGAGGAAAGCAAAACCTTAACTGAACTATGACAAACTCAGCCCTGCACTTCGCT 1089
 DB 1076 GGGGCACTCCAGAGAACCGGCCAGCCATGATTTACGACAGCTGAGCCGCTGCTCCGAT 1135
 QY 1090 ACTACTATGACAAAAATATTTATGACTTAAAGTTTCAATGCTTAAAGCTTATGCTTAAAGTTT 1149
 DB 1136 ACTATTATGAGAAAGGCATCATGCGAGAGTGGCTGAGCGTTTACGTGTACAAAGTTTG 1195

Search completed: May 27, 2004, 19:27:30
 Job time : 147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2004, 17:48:07 ; Search time 725 Seconds
(without alignments)
9078.980 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447

Sequence: 1 gaattccgcgaacgaataat.....gaaagaagcccaagaaaa 1447

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/FCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1447	100.0	1447	9	US-09-902-772-1
2	1356	93.7	1528	9	US-09-902-772-3
3	921	63.6	1389	13	US-10-211-462-41
4	921	63.6	1389	13	US-10-087-192-1127
5	921	63.6	3166	15	US-10-205-823-98
6	921	63.6	3166	15	US-10-021-660-26
7	763.2	52.7	1297	13	US-10-087-192-1124
8	570.2	39.4	1359	12	US-10-052-482-195
9	570.2	39.4	1729	12	US-10-052-482-194
10	562.2	38.9	1359	12	US-10-052-482-198
11	562.2	38.9	2957	12	US-10-052-482-197
12	562.2	38.9	2957	13	US-10-342-887-710
13	562.2	38.9	2957	13	US-10-172-118-710
14	562.2	38.9	2957	15	US-10-007-926A-295

C	15	387	26.7	567	9	US-09-864-761-18410	A	Sequence 18410, A
	16	387	26.7	212231	13	US-10-087-192-1126	Ap	Sequence 1126, Ap
	17	386	26.7	69090	13	US-10-087-192-1123	Ap	Sequence 1123, Ap
C	18	354.6	24.5	473	9	US-09-864-761-20472	A	Sequence 20472, A
C	19	353.6	24.4	472	15	US-10-029-386-25224	A	Sequence 25224, A
	20	321.4	22.2	72732	12	US-10-052-482-193	Ap	Sequence 193, Ap
	21	320.4	22.1	73725	12	US-10-052-482-196	Ap	Sequence 196, Ap
	22	307	21.2	467	10	US-09-918-995-23356	A	Sequence 23356, A
	23	276.6	19.1	549	9	US-09-923-779-54	Appl	Sequence 54, Appl
C	24	255	17.6	420	9	US-09-864-761-3705	Ap	Sequence 3705, Ap
C	25	255	17.6	454	9	US-09-864-761-1653	Ap	Sequence 1653, Ap
	26	180	12.4	1752	9	US-09-850-799-1	Appl	Sequence 1, Appl
	27	180	12.4	1752	15	US-10-027-859-1	Appl	Sequence 1, Appl
C	28	167.8	11.6	553	9	US-09-864-761-7427	Ap	Sequence 7427, Ap
	29	166.2	11.5	380	15	US-10-263-828-41	Appl	Sequence 41, Appl
C	30	163.2	11.3	348	9	US-09-864-761-29364	A	Sequence 29364, A
C	31	163.2	11.3	477	9	US-09-864-761-5035	Ap	Sequence 5035, Ap
C	32	163.2	11.3	519	9	US-09-864-761-12798	A	Sequence 12798, A
C	33	162.2	11.2	225	9	US-09-864-761-21761	A	Sequence 21761, A
C	34	153.8	10.6	530	15	US-10-029-386-11517	A	Sequence 11517, A
	35	148.4	10.3	2532	15	US-10-037-270-869	Ap	Sequence 869, Ap
	36	148.4	10.3	2532	16	US-10-117-722-869	Ap	Sequence 869, Ap
	37	131.8	9.1	1496	13	US-10-087-192-344	Ap	Sequence 344, Ap
	38	125.4	8.7	4991	12	US-10-152-319A-1502	Ap	Sequence 1502, Ap
	39	124.6	8.6	1450	16	US-10-191-997-126	Ap	Sequence 126, Ap
	40	123	8.5	1602	13	US-10-087-192-347	Ap	Sequence 347, Ap
	41	122.8	8.5	416	10	US-09-918-995-8472	Ap	Sequence 8472, Ap
C	42	122.4	8.5	165	9	US-09-864-761-23294	A	Sequence 23294, A
	43	122.2	8.4	441	15	US-10-210-120-41	Appl	Sequence 41, Appl
	44	122.2	8.4	473	10	US-09-918-995-27904	A	Sequence 27904, A
	45	122.2	8.4	1884	9	US-09-925-300-420	Ap	Sequence 420, Ap

ALIGNMENTS

RESULT 1

US-09-902-772-1
; Sequence 1, Application US/09902772
; Patent No. US20020164739A1
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwanoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; TITLE OF INVENTION: the Proteins
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/09/902,772
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US/08/878,177
; PRIOR FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA
US-09-902-772-1

Query Match	100.0%	Score 1447	DB 9	Length 1447
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1447	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	GAATTCGCGAAGCAATAATATTATTAGCAATATTAGCGATCAATAATCTTGATCACA	60	
Db	1	GAATTCGCGAAGCAATAATATTATTAGCAATATTAGCGATCAATAATCTTGATCACA	60	
Qy	61	TTATGCGAAGCACTATTATAGGAGCATTATCAGTGGTGAAGACCACTCTTGTGTTG	120	
Db	61	TTATGCGAAGCACTATTATAGGAGCATTATCAGTGGTGAAGACCACTCTTGTGTTG	120	
Qy	121	AGTGTGCTAGGATCGCCCACTTTCGAAAGACAGAAATGACAGCCTCTTCCAGTG	180	
Db	121	AGTGTGCTAGGATCGCCCACTTTCGAAAGACAGAAATGACAGCCTCTTCCAGTG	180	

Db 573 CGCGACATCCTTCTCTCACATCTCCACTACCTCAGAGAGACTCCTCTTCCACATTTGAC 632
Qy 665 TTTTATTTTCCAAATACATCAGTTTACCAGAGAGCAACGACAGAGATTAACACAGGCC 724
Db 633 TTCAGATGATGTTGATAAGCTTACAAAACCTCCAGGTTATGATGCTGAACAC 632
Qy 725 AGATTACCTTATGAGCAAGCGAGAGATCAGCTGAGAGTCCACAGCCATCCCA--C 781
Db 693 AGATTACCATATGAGCCCCCAGAGATCAGCTGAGAGTCCAGCGGTCCAGGCCACCCACGCC 752
Qy 782 TCAGTCAAAGCTACCCCAACCATCATCTTCAACAGTGCACAAAACAGAGACAGCGTCC 841
Db 753 CCAGTGAAGCTGCTCAACCATCTCTCTTCCAGAGTGCACAAAACAGAGACAGCGTCC 812
Qy 842 TCAGTTAGATCCTTATCAGATCTTGGACCGACGAGCGCGTCTTCCAAATCCAGGAG 901
Db 813 TCAGTTAGATCCTTATCAGATCTTGGACCAACAGTAGCGCGCTTCCAAATCCAGGAG 872
Qy 902 TGGGAGATACAGCTATGCGAGTCTTCTACTGAGCTTCTGCGACAGCTCCAACTCAA 961
Db 873 TGGCCAGATCAGCTTGGAGTCTTCTCTGAGCTCTGTCGAGACGCTCCAACTCAG 932
Qy 962 CTGCATACCTGGGAGGACAAATGGGAGTTCAGAGTACAGACCTGATGAAGTGGC 1021
Db 933 CTGCATACCTGGGAGGACAAACGAGGAGTTCAGAGTACAGCTCCGAGGTTGGC 992
Qy 1022 TGGGCTTGGGAGAGAGAAAGCAACCTAACTGNACTATGACAACTCAGACCTGTC 1081
Db 993 CCGCGCTGGGAGAGCGAGAGAAACCAACCAATGAACTACGATGAAGTTCAGCGCGC 1052
Qy 1082 ACTTCGCTACTATGACAAAATATTTATGACTAAAGTTTCATGGTAAAGCTATGCTTA 1141
Db 1053 CTTCTGTTACTATGACAAAGACATCATGACCAAGTTCATGGGAGGCTACGCTTA 1112
Qy 1142 CAATTTGATTTCCAGGAATCGCTCAGGCTCCTCAGGCTCCTCAGCTCCTCAGATCATCAT 1201
Db 1113 CAAGTTCGACTTCCAGCGGATCGCCAGGCTCCTCAGGCTCCTCAGGCTCCTCAGTCTCT 1172
Qy 1202 GTACAAATACCATCAGACCTCCCTACATGAGTTCTTACCATGACACACCCAGAGAT 1261
Db 1173 GTACAAATACCATCAGACCTCCCTACATGAGTTCTTACCATGACACACCCAGAGAT 1232
Qy 1262 GAATTTGAGTCCCATCCCTCCTGCTTTCGCGTAACTCATCAGCTTCTTGTGTC 1321
Db 1233 GAATTTGAGTCCCATCCCTCCTGCTTTCGCGTAACTCATCAGCTTCTTGTGTC 1292
Qy 1322 CCCTAATCCATCGAATTCACCAACTGAGGAGTCTTACCCTCAATACCGGCTGCCAGC 1381
Db 1293 CCCTAATCCATCGAATTCACCAACTGAGGAGTCTTACCCTCAATACCGGCTGCCAGC 1352
Qy 1382 TGCTCATATGCTTCCCATCTTGGACCTACTACTAA 1418
Db 1353 CAGCCATATGCTTCTCATCTGGGACCTTACTACTAA 1389

RESULT 4

US-10-087-192-1127
; Sequence 1127, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: CANCER
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PENDING FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1127
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1127

Query Match 63.6%; Score 921; DB 13; Length 1389;
Best Local Similarity 81.1%; Pred. No. 1.4e-294;
Matches 1100; Conservative 0; Mismatches 245; Indels 12; Gaps 2;

Qy 74 TATTAAGGAAGCATATCAGTGTGAGTGAAGACAGTCTTGTGTTGAGTGTGCTACGG 133
Db 33 TATCAGGAGAGCTTATCAGTTGTGAGTGAAGACAGTCTTGTGTTGAGTGTGCTACGG 92
Qy 134 ATGCGCCCACTTGCAAAAGACAGAAATGACAGCTCTCTTCCAGTGAATATGGGCAAC 193
Db 93 AACGCCACACCTTGGTAAAGACAGAGATGACCGCTCTCTCCAGCGACTATGGACAGC 152
Qy 194 ATCAAGATGAGCGCGGTTCCAGCAGAGCTGTTTATCAGACGCCCGCCAGAGT 253
Db 153 TTCCAGATGAGCCCAAGCGCTCTCAGCAGATTTGGTGTCTCAACCCCAAGCAGGCT 212
Qy 254 TACCATTAAAGATGAGTGTAAACCAACAGGTTTAAATGGTCAAGAAATTCACCTGATGA 313
Db 213 CACCATCAAAATGGAATGTAACTTAGCCAGGTGAATGGCTCAAGGAACTCTCTGATGA 272
Qy 314 CTGCAGCTGGCAAAAGAGGAGAAATGGTTAGCAGTTTCAAGCAATGTTGGGATGACTA 373
Db 273 ATGCAAGTGGGCCAAAGCGGGAAGATGGTGGGAGCCAGACACCGTTGGGATGACTA 332
Qy 374 TGAAGCTACATGGAAGAGAGACATATTCGCGCTCCAAATATGACAAACCAATGAACGAG 433
Db 333 CGGAGCTTACATGGAGGAGAGACATGATCCACCCCAACATGACCAAGAGGCGAG 392
Qy 434 AGTTATGTGCCAGAGATCTTACGTTATGGAGACAGACCATGTACGCGAGTGGCTGA 493
Db 393 AGTTATGTGCCAGAGATCTTACGTTATGGAGTACAGACCATGTGCGGAGTGGCTGA 452
Qy 494 GTGGGAGTGAAGGAGTATGTTTCCAGACCTGACATCTTGTGTTCCAGAACTTGA 553
Db 453 GTGGGAGTGAAGGAGTATGTTTCCAGACCTGACATCTTGTGTTCCAGAACTTGA 512
Qy 554 TGGGAAGAGTGTGTAATGACCAAGATGACTTCCAGAGACTCAGCGGAGCTATAA 613
Db 513 TGGGAAGAGTGTGTAATGACCAAGATGACTTCCAGAGGCTCAGCGGAGCTATAA 572
Qy 614 CGCAGATATCTCTCTGTCACACTACCTACCTCAGAGAGA-----GAGGAGCCAC 664
Db 573 CGCGGACATCTCTCTCTCAGATCTCCACTACCTCAGAGAGACTCCTCTTCCACATTTGAC 632
Qy 665 TTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAGTCTTGGAAATCCAGGAG 724
Db 633 TTCAGATGATGTTGATAAAGCTTACAAAACCTCTCCAGGTTTAAATGCAATGTAGAAACAC 692
Qy 725 AGATTTTACCTTATGAGCAAGCGAGAGATCAGCGTGGAGAGTCAAGCATCTCCCA--C 781
Db 693 AGATTTTACCTTATGAGCCCCCAGAGATCAGCTTGGAGCGTCAAGCGGACCCCAAGCC 752
Qy 782 TCAGTCAAAGCTACCCCAACATCTTTCAGAGTGGCCCAAAACAGAGACAGCGTCC 841
Db 753 CCAGTCAAAGCTGCTCAACCATCTCTCTTCCAGAGTGGCCCAAAACAGAGACAGCGTCC 812
Qy 842 TCAGTTAGATCCTTATCAGATCTTGGACCGACAGCGCTCTTGGAAATCCAGGAG 901
Db 813 TCAGTTAGATCCTTATCAGATCTTGGACCAACAGTAGCGCGCTTGGAAATCCAGGAG 872
Qy 902 TGGGAGATACAGCTATGCGAGTCTTCTACTGAGCTTCTGTCGAGCAGCTCCAACTCAA 961
Db 873 TGGCCAGATCCAGCTTGGAGTCTTCTCTGAGCTCCTGTCGAGAGCTCCAACTCCAG 932
Qy 962 CTGCATCCTGGGAGGACAAATGGGAGTTCAGATGACAGACCTGATGAAGTGGC 1021
Db 933 CTGCATCCTGGGAGGACCAACAGGAGTTCAGATGACCGATCCCGAGGAGTGGC 992

QY 1022 TCGGCGTTGGGAGAGAGAAAGCAAACTTAACATGAACATATGACAAACTCAGCCGTC 1081
DB 993 CCGGCGCTGGGAGAGCGGAAGAGCAAACTTAACATGAACATACGATAAGCTCAGCCGCGC 1052
QY 1082 ACTTCGCTACTACTATGACAAATAATATTAGTAAAGTTCATGTTAAACGCTATGCCTA 1141
DB 1053 CTTCCGTTACTACTATGACAAAGACATCATGACCAAGTCCATGGGAAGCGCTAGCCCTA 1112
QY 1142 CAAATTTGATTCCACGGAATCGCTCAGGCCCTCCAGCCTCACCTCCAGAAATCATCCAT 1201
DB 1113 CAAATTCGACTTCCACGGGATCGCCAGGCCCTCCAGGCCCTCCAGGCCCTCCAGGCCCT 1172
QY 1202 GTACAAATACCCATCAGACCTCCCTTACATGATGTTCTTACCATGACACACCCCGAGAGAT 1261
DB 1173 GTACAAAGTACCCCTCAGACCTCCCTTACATGATGTTCTTACCATGACACACCCCGAGAGAT 1232
QY 1262 GAACTTTGTAGTCTCCCATCCCTTGTCTTGGCCCTTAACCTCATCCAGCTTTTGTGTC 1321
DB 1233 GAACTTTGTGGCGCCCACTCCAGCCCTCCCGTGACATCTTCCAGTCTTTTGTGTC 1292
QY 1322 CCTTAATCCTACTGGAATTCACCACTGGAGGATCTACCCCAATACCAAGCTGCCAGC 1381
DB 1293 CCAAAACCCATCTGGAATTCACCACTGGGATATATACCCCAACACACTAGGCTCCCCAC 1352
QY 1382 TGCTCATATGCTTCCCATCTTGGCACTACTACTAA 1418
DB 1353 CAGCCATATGCTTCTCATCTGGGCACCTACTACTAA 1389

RESULT 5

US-10-205-823-98
; Sequence 98, Application US/10205823
; Publication No. US20030108963M1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 3166
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-98

Query Match 63.6%; Score 921; DB 15; Length 3166;
Best Local Similarity 81.1%; Pred. No. 2.5e-294;
Matches 1100; Conservative 0; Mismatches 245; Indels 12; Gaps 2;

QY 74 TATTAAGGAAGCATTTATCAAGTGGTGAAGACCAAGTCTCTTGTGTTGAGTGTCCCTACGG 133
DB 289 TATCAAGGAAGCCTTATCAAGTGGTGAAGACCAAGTCTCTTGTGTTGAGTGTCCCTACGG 348
QY 134 ATGCCCCCCTTGCRAAGACAGAAATGACAGCTCTCTTCCAGTGAATATATGGGCAAAAC 193
DB 349 AAGCCACACCTTGGCTAAGACAGAGATGACCGGCTCTCTTCCAGGACATATGGACAGAC 408
QY 194 ATCAAGATGAGCCCGCGCTTCCCGAGGAGTCTGTTATCACAGCCCCCGCCAGAGT 253
DB 409 TTCAAGATGAGCCCGCGCTCTCTCAGCAGGATTTGGCTGTCTCAACCCCGCCAGGAGT 468
QY 254 TACCATTAAGATGAGCTGTAAACCAACAGGTTAATGGGTCAAGGAATTCACCTGATGA 313
DB 469 CACCATCAAAATGGAATGTAAACCTTAGCCAGGTTGAATGGCTCAAGGAATCTCTCTGATGA 528
QY 314 CTGACGCTGGCAAAAGGAGGAAATGGTTAGCAGTTTACAGATTCAGACAATGTTGGGATGAACCTA 373
DB 529 ATGAGTGTGGCAAAAGGCGGGAAGATGTTGGGACCCCGACACACCTTTGGGATGAACCTA 588
QY 374 TGAAGCTACATGGAAGAGAGACATATTCGGCTTCAAAATATGACAAACCAATGAAAG 433
DB 589 CGCAGCTACATGGAAGAGAGACATGCCCCCAAAACATGACCAACGAGCGCAG 648
QY 434 AGTTATTGTGCCAGCAGATCTCTACGTTATGGAGACAGACCATGTACGCGAGTGGGTGGA 493
DB 649 AGTTATCGTGGCAGCAGATCTCTACGCTATGGAGTACAGACCATGTGCGCAGTGGCTGGA 708
QY 494 GTGGCAGTGAAGAGTATGCTTCCAGACGTGGACATCTTGTGTTCCAGAACATGTA 553
DB 709 GTGGCGGTTGAAAGAAATATGGCCTTCCAGACGTTCAACATCTTGTATTCAGAACATCGA 768
QY 554 TGGGAAGAGTGTGTAAATATGACCAAGATGACTTCCAGAGACTCAGCCGAGCTATAA 613
DB 769 TGGGAAGAGTGTGTGAAGATGACCAAGACGACTTCCAGAGCTCACCCCGAGCTACAA 828
QY 614 CGCAGATATCTCTCTGTCACCTACACTACTCTCAGAGAGA-----GAGAGCCAC 664
DB 829 CGCCGACATCTCTCTCACATCTCCACTACTCTCAGAGAGACTCTCTCTTCCACATTTGAC 888
QY 665 TTTTATTTTCCAAATACATCAGTTTACCAGAGCAACGCAAGAAATAACACAAGGCC 724
DB 889 TCCAGATGATGTTGATAAGCCCTTACAAACTCTCCAGGTTAATGCTATGCTAGAACAC 948
QY 725 AGATTTACCTTTATGACCAAGCAGAGAGTACGCTGGAGAGTACAGCCATCCCA---C 781
DB 949 AGATTTACCATATGAGCCCCCAGGAGATCAGCTGGACCGGTACGCGCACCCACGCC 1008
QY 782 TCAGTCAAAAGCTACCCACCATCATCTTCAACAGTGGCCCAACAGACAGACCCAGGTC 841
DB 1009 CCAGTCGAAAGCTGCTCAACCATCTCTTCCAGTGGCCCAAACTGAAGACCCAGGTC 1068
QY 842 TCAGTTAGATCTTTATCAGATCTTGGACCGACAGCAGCCGCTTTGCAAAATCCAGGAG 901
DB 1069 TCAGTTAGATCTTTATCAGATCTTGGACCAACAAGTAGCCGCTTTGCAAAATCCAGGAG 1128
QY 902 TGGGCGATACAGCTATGCGAGTTCTTCTGAGCTTCTGTGGAGAGCTCCCACTCCAA 961
DB 1129 TGGCCAGATCCAGCTTGGCAGTCTCTCTGGAGCTCTCTGTGGAGAGCTCCCACTCCAG 1188
QY 962 CTGCATCACCTGGGAGGCGACAAATGGGAGTTCAAGATGACAGACCCCTGATGAAGTGGC 1021
DB 1189 CTGCATCACCTGGGAGGCGACCAACCGGAGTTCAAGATGACAGATCCCGCAGGTTGGC 1248
QY 1022 TCGGCGTTGGGAGAGAGAAAGCAACCTTAACATGAACTATGACAAACTCAGCGCTGC 1081
DB 1249 CCGGCGCTGGGAGAGCGGAAGCAAAACCAACATGAATAGCTACGAGCTACGCGCGC 1308
QY 1082 ACTTCGCTACTACTATGACAAATAATATTATGCTAAAGTTTATGGTAAACGCTATGCCTA 1141
DB 1309 CTTCCGTTACTACTATGACAAAGAACTATGACCAAGTCCATGGAGCGCTACGCTA 1368


```

US-10-087-192-1124
; Sequence 1124, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1124
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-1124

Query Match      52.7%; Score 763.2; DB 13; Length 1297;
Best Local Similarity 75.3%; Pred. No. 4,1e-242;
Matches 1008; Conservative 0; Mismatches 288; Indels 42; Gaps 3;

QY 80 GGAAGCATTTACGTGTGAGTGAAGACCACTCTTTGTTGAGTGTGCTTACGGATGCC 139
DB 1 GGAGGCCCTTGTGAGTGTGAGGAGGACCACTCACTATTGAGTGTGCTTACGGAGGCC 60

QY 140 CCACCTTGCAGAGACAGAAATGACAGCTCTCTTCCAGTGAATATGGGCAACATCAGA 199
DB 61 ACACCTGGCTAAGACAGAGATGACCGCATCTCTTCCAGTGAATATGGGCAACATCAA 120

QY 200 GATGAGCCCGCGCTTCCACAGAGAGTGTGTTATACAGCCCCCGCCAGAGATTACCAT 259
DB 121 GATGAGTCCCAGAGTCCCTCAGCAGGACTGCTGTCTCAAGCCCCAGCCAGGATCACCAT 180

QY 260 TAAGATGAGTGTAAACCAACAGGTTAATGGTCAAGGAATTCACCTGATGACTGCAG 319
DB 181 CAAGATGAGTGCACACCTAGTCAAGGTGAATGGTTCAGGAACATCACTGATGAGTGCAG 240

QY 320 CGTGGCAAAAGAGGAGAAATGGTTAGCAGTTTCAGACAAATGTTGGGATGAACATATGGAAG 379
DB 241 TGTGAACAAAGGTGGAGAGTGTGGGAGCCCGGATGACTGTGGGATGAGCTACGGCAG 300

QY 380 CTACATGGAAGAGAGATATATCCGCTCCAAATATGACAAACCAATGAACGAGAGTTAT 439
DB 301 CTACATGGAAGAGAGATGATGGCGCTCCCAATATGACCAACAAATGAGCGCAGAGTGAT 360

QY 440 TGTGCCAGCAGATCCTACGTTATGGAGCAGACAGCATGTACGGCAGTGGCTGAGTGGGC 499
DB 361 CGTCCCTGCAGATCCTACTCTGTGAGACACAGACCATGTCCGACAGTGGCTGAGTGGGC 420

QY 500 AGTGAAGAGATGGTCTTCAGACGTTGGACATCTTGTGTTCCAGAACATTTGATGGAA 559
DB 421 GGTGAAAGAAATATGGCTCTCGATGTGGAGCTTACTATTTTCAGAAATATCGATGGAA 480

QY 560 AGAGTGTGTAATGACCAAGATGACTTCCAGAGACTCAAGCCGAGCTATAACGCAGA 619
DB 481 GGAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGCTCAGCCGAGCTACATGCCGA 540

QY 620 TATCCTCTGTACACCTACACTACCTCAGAGAGAGAGGCCACTTTTATTTTCCAAA 679
DB 541 CATTTCTCTCTCACATCTCCACTACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 584

QY 680 TACATCAGTTTACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 739
DB 585 CACATCTGACTTCCGATGACCTTGATAGGCTTTTACAAA-----CTCTCCAGGTTAAT 639

QY 740 GCAAGCGAGGAGATCAGCGTGGAGAGTTCACAGCCATCCCACTCAGTCAAAAGATCCCA 799

```

RESULT 8

```

US-10-052-482-195
; Sequence 195, Application US/10052482
; Publication No. US2004007226A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 195
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-052-482-195

```

```

Query Match      39.4%; Score 570.2; DB 12; Length 1359;
Best Local Similarity 66.8%; Pred. No. 6.3e-178;
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;

```

QY 63 ATGGCAAGCACTATTAAAGGAGCATTATCAGTGTGAGTGAAGACCAAGTCTCTTTTGGAG 122
 Db |||||
 QY 1 ATGGAGGGAGCTATTAAAGAGGCTCTGTCTGTGTGAGTGAAGTCAAGTCCCTTTTGGAT 60
 Db |||||
 QY 123 TGTGCTACGGATCG---CCCAACCTTGCAGAGAGCAGAAATGACAGAGCTCTCTTCCAGT 179
 Db |||||
 QY 61 TCAGCATACGGAGCGGAGCCCATCTCTCCCAAGGCAGATATGACTGTCTCGGAGGCTCT 120
 Db |||||
 QY 180 GAAATATGGGCAACATCAAGATGAGCCCGCGGTTCACAGAGAGTGGTTATACAG 239
 Db |||||
 QY 121 GACTACGGGAGCGCCCAAAAATCAACCCCTGCGCACCGCAGCAGGATGGATCAACAG 180
 Db |||||
 QY 240 CCCCCGGCCAGAGTTACCAATTAAGATGAGTGTAAACCCCAACACAGGTTAATGGGTCAAGG 299
 Db |||||
 QY 181 CC-----AGTCAGAGTCAATGTCAAGCGGAGTATGACCATGAATGGATCCAGG 231
 Db |||||
 QY 300 AATTCACCTGATGCTGAGCGGTGCAAAAAGAGGAGGAAATGGTTAGAGTTACAGCAAT 359
 Db |||||
 QY 232 GAGTCTCCGGTGGAGTGCAGTGTGAGCAATGTAAACAGTGTGGGCGGAGGCGAAGCC 291
 Db |||||
 QY 360 GTTGGGATGAACTATGGAGCTACATGGAAGAGAGCA---TATTCCGCTCCCAATATG 416
 Db |||||
 QY 292 AACCCCATGAACATAAATAGCTACATGATGAGAGAAACGGCCCTCTCTCCCAATG 351
 Db |||||
 QY 417 ACAACCAATGACGAGAGTTATTGTGCCAGCAGATCTCTACGTTATGAGCAGACCAAT 476
 Db |||||
 QY 352 ACCACCAAGCAAGAGAGTCTATTGTGCTGCAGACCCCACTGTGCACAGAGGAGC 411
 Db |||||
 QY 477 GTACGGCAGTGGCTGAGTGGCAGTGAAGAGTATGCTCTCCAGAGTGGACATCTTG 536
 Db |||||
 QY 412 GTTCAGCAGTGGCTGGAGTGGCTATAAAGAAATACGGAATGTAGGAGATTGACATCTCC 471
 Db |||||
 QY 537 TTGTTCCAGAACTATGATGGGAAAGAGTTGTGTAAATGACCAAGATGATCTCCAGAG 596
 Db |||||
 QY 472 TTCTCCAGAACATGATGGCAAGAAATGTGTAAATGAACAAGGAGGACTTCTCCGA 531
 Db |||||
 QY 597 CTCAGCGCAGCTATAAGCAGATATCTCTGTGACACCTACACTACCTCAGAGAGAG 656
 Db |||||
 QY 532 GCCACCTCCGCTACAACACAGAAAGTGTGTGTGCACTCTCAGTTACCTCAGGGAAG- 590
 Db |||||
 QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAAATAACA 716
 Db |||||
 QY 591 -----TTCACTGCTGCTCTATAACACAACTCCCATACAGACCACTCTCAGACTGAT 645
 Db |||||
 QY 717 ACAAGCCAGATTTTACCTTATGAGCAACGAGGAGATCAGCGTGAGAGTCAACA---GC 773
 Db |||||
 QY 646 GTCAAGGAAGACCTTCTTATGACTCTCTCAGGAGAGGAGCATGGAACAATAATGAAC 705
 Db |||||
 QY 774 CATCCCACTCAGTCAAAAGCTACCAACCATCTCTTCAACAGTCCCAACAGAGAGAC 833
 Db |||||
 QY 706 TCTGGCTCAACAAAGTCTCTCTCTGGAGATCAAGACCATGGGAGAGACACTGAG 765
 Db |||||
 QY 834 CAGCGTCTCAGTTAGATCTTATCAGATTCTTGGACCGACAGCAGCGCTTTCGCAAT 893
 Db |||||
 QY 766 CAGCGCCCCAGCAGATCTTATCAGATCTTGGGCGCAACAGCAGCGGCTAGCAAC 825
 Db |||||
 QY 894 CAGGAGTGGCAGATACAGATATGAGCTTCTCTGAGAGTCTCTGCGAGAGCTCC 953
 Db |||||
 QY 826 CTTGGAGTGGCAGATCAGCTGTGGCAGTTCTCTTGGAACTACTGTCCGAGAGGCGCC 885
 Db |||||
 QY 954 TAATCCAACTGATCACTCTGGAGGCGCAAAATGGGGAGTTCAAGATGACAGACCTCTGAT 1013
 Db |||||
 QY 886 AACCCAGCTGTATCACTCTGGAGGCGCAACAGGAGTTCAAAATGACGAGCCCTGAT 945
 Db |||||
 QY 1014 GAAGTGGCTGGCGTTGGGAGAGAGGAAAGCAACCTAACATGAATATGACAACTC 1073
 Db |||||
 QY 946 GAGTGGCGAGCGCTGGGAGAGCGGAGAGCAAGCCCAACATGAATATGACAACTG 1005
 Db |||||
 QY 1074 AGCCGTCAGCTTCGCTACTACTATGACAAAATATTATGATTAAGTTTCATGGTAAACGC 1133
 Db |||||
 QY 1006 AGCCGGGCCCTCCGATACTACTATGACAAACATTTATGACCAAGTCTATGGCAAGG 1065
 Db |||||

QY 1134 TATGCTCAAAATTTGATTTCCAGGAATCGTCCAGCCCTCAGCCTCACCTCCAGAA 1193
 Db |||||
 QY 1066 TATGCTCAAAATTTGATTTCCAGGAATCGTCCAGCCCTCAGCCTCAACAGAG 1125
 Db |||||
 QY 1194 TCATCCATACAAATACCAATACCACTCCCTACATGAGTTCTTACCAATCAACACCC 1253
 Db |||||
 QY 1126 ACATCCATGACAAAGTATCCCTCTGATATCTCTACATGCTCTTCTTACCATCCATCAA 1185
 Db |||||
 QY 1254 CAGAAGATGAATTTGATCTCCCTCCCTCCCTCTCTTTCCTGCTTTCCTTCCAGCTTT 1313
 Db |||||
 QY 1186 CAGAAGTGAATTTGATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1245
 Db |||||
 QY 1314 TTTGCTGCCCTTAATCCATACCTGGAATTCACCAATCGGAGGATCTTACCCCAATACAGG 1373
 Db |||||
 QY 1246 TTTGGAGCAGCATCAATATCTGAGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1305
 Db |||||
 QY 1374 CTGCCAGCTGCTCAT- - - - -TGCTTCCCTCTTGGCACTTACTACTA 1417
 Db |||||
 QY 1306 GTCCCCCGCCATCTTAACACCCACGTCGCTTACACTTAGGAGCTACTACTA 1358
 Db |||||

RESULT 9
 US-10-052-482-194
 ; Sequence 194, Application US/10052482
 ; Publication No. US20040072264A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Engelhard, Eric
 ; APPLICANT: Morris, David
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
 ; FILE REFERENCE: A-71087/RMS/DCF
 ; CURRENT APPLICATION NUMBER: US/10/052,482
 ; CURRENT FILING DATE: 2002-08-15
 ; PRIOR APPLICATION NUMBER: US 09/747,377
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: US 09/798,586
 ; PRIOR FILING DATE: 2001-03-02
 ; NUMBER OF SEQ ID NOS: 241
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 194
 ; LENGTH: 1729
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-052-482-194

Query Match 39.4%; Score 570.2; DB 12; Length 1729;
 Best Local Similarity 66.8%; Pred. No. 7.4e-178; Indels 33; Gaps 6;
 Matches 917; Conservative 0; Mismatches 423;
 QY 63 ATGGCAAGCACTATTAAAGGAGCATTATCAGTGTGAGTGAAGACCAAGTCTCTTTTGGAG 122
 Db |||||
 QY 245 ATGGACGGGACTATTAAAGGAGGCTCTGTCTGTGTGAGTGAAGTCAAGTCCCTTTTGGAT 304
 Db |||||
 QY 123 TGTGCTACGGATCG---CCCAACCTTGCAGAGAGAGAAATGACAGCCTCTCTTCCAGT 179
 Db |||||
 QY 305 TCAGCATACGGAGCGGAGCCCATCTCCCAAGGAGAGATATGACTGCTTCGGGGAGTCT 364
 Db |||||
 QY 180 GAATATGGGCAACATCAAGATGAGCCGCGTTCCCAAGCAGGACTGGTTATCACAG 239
 Db |||||
 QY 365 GACTACGGGAGCGCCCAAAATCAACCCCTGCGCACCGCAGCAGGAGTGGATCAACAG 424
 Db |||||
 QY 240 CCCCCGGCAGAGTTACCAATTAAGATGAGTGTAAACCAACAGGTTAATGGTCAAGG 299
 Db |||||
 QY 425 CC-----AGTCAGAGTCAATGTCAAGCGGAGTATGACCAATGAATGATCCAGG 475
 Db |||||
 QY 300 AATTCACCTGATGACTGACGCTGGCAAGAGGAGGAAATGTTAGCAGTTTCAACAAT 359
 Db |||||
 QY 476 GAGTCTCCGGTGGAGTGTGAGTGTGAGCAATGTAAACAGCTGGTGGCGGAGCGAAGCC 535
 Db |||||
 QY 360 GTTGGGATGAACCTATGGAAGCTACATGGAAGAGAGCA---TATTCCGCTTCAAAATG 416
 Db |||||
 QY 536 AACCCCATGAACCTATAATAGTACATGATGAGAGAGCGGCCCTCTCTCCCAACATG 595
 Db |||||
 QY 417 ACAACCAATGAAGAGAGTATTATTGGCCAGCAGATCTCTACGTTATGAGCAGACCAAT 476
 Db |||||

Db 596 ACCACCAAGACGAGAGATCTTGTGCTGCAGACCCACACCTGTGTGACACAGGAGCAC 655
Qy 477 GTACGGAGAGTGGTGGAGTGGAGAGTATGCTTCCAGACGTTGGAGCATCTTG 536
Db 656 GTTCAGACATGGTGGAGTGGAGTATGAGGAATACGAGATGATGGAGATGACACTTCC 715
Qy 537 TTGTTCCAGAACTTGTATGGGAGAGAGTGTGTAATGACCAAGATGACTTCCAGAGA 596
Db 716 TTCTTCCAGAACATGGATGGCAAGGAATGTGTAAATGAACAAGGAGGACTTCTCCGA 775
Qy 597 CTCAGCGGAGCTATAAGCAGATATCTCTCTGTCTACACCTACACTACCTCAGAGAGA 656
Db 776 GCCACCTCCGCTCAACACAGAGATGCTGTGTGCGACCTCAGTTACCTCAGGGAAG- 834
Qy 657 GGAGCCATTTTATTTTCCAAATACATCAGTTTACCAGAGCAACGCAAGAAATAACA 716
Db 835 -----TTCACTGTGGCTATAACACACCTCCATACAGACAGTCTCAGACTGAT 889
Qy 717 ACAAGGCCAGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTACA---GC 773
Db 890 GTCAAGGAAGACCTTCTTATGACTCTGTCTAGGAGAGGAGCATGGAACAATAATATGAC 949
Qy 774 CATCCCACTAGTCAAAAGCTACCAACCATCTTCAAGTTCAGAGTCCCAAAACAGAGAC 833
Db 950 TCTGGCTCAACAAAGTCTCTCTCTGGAGGATCAAGACCATGGGCAAGAACACTGAG 1009
Qy 834 CAGCGTCTCAGTTAGATCTTATCAGATTTCTGGACCGACAGCAGCCGCTTTGCAAT 893
Db 1010 CAGCGCCCGACGAGATCTTATCAGATCTCTGGGCGCCACACGACGCGCTAGCAAC 1069
Qy 894 CGAGGAGTGGGAGATACAGCTATGGAGTCTCTACTGGAGCTTCTGTGGAGAGTCC 953
Db 1070 CTTGGAGTGGGAGATCCAGCTGTGGAGTTTCTCTGGAACTACTGTCCGACAGCGCC 1129
Qy 954 AACTCCAACTGATCAGCTGGGAGGCAAAATGGGAGTTCAAGATGACAGACCTGAT 1013
Db 1130 AACGCCAGCTGATCAGCTGGGAGGAGCAACCGGGAGTTCAAAATGACGGACCTGAT 1189
Qy 1014 GAAGTGGTGGGAGAGAGAAAGCAAACTTAATGAACTATGACAACTC 1073
Db 1190 GAGTGGCCAGCGCTGGGAGAGCGGAAGAGCAAGCCCAACATGAATTTATGACAGCTG 1249
Qy 1074 AGCGTGACATTCGCTACTACTATGACAAATAATATGACTAAAGTTCTAGTAAACGC 1133
Db 1250 AGCGGGCCCTCCGATACTATGACAAACAACTATGACAAAGTGCATGGCAAGG 1309
Qy 1134 TATGCTTACAAATTTGATTTCCAGGAATCGCTCAGCGCCCTCCAGCTCAGCTTCCAGAA 1193
Db 1310 TATGCTTACAAATTTGATTTCCATGGCATTCGCCAGCGCTCCAGCCACATCCACAGAG 1369
Qy 1194 TCATCCATGTAACAAATACCAATCAGACCTCCCTACATGATTTCTTACCATGACACCCC 1253
Db 1370 ACATCCATGTAACAGTATCCCTCTGATATCTCTTACATGCTTCTTACCATGCCCATCA 1429
Qy 1254 CAGAGATGAACTTTGTAGCTCCCATCCCTGCTTTGCGCGCTAACTCATCCAGCTTT 1313
Db 1430 CAGAGAGTGAATTTGTCCGCTCTACCCATCCCTCCATGCTGCTCCTCCAGCTTC 1489
Qy 1314 TTGTGCTCCCTTAATCCATATGAGAAATTCACCACTGAGGAGATCTACCCCAATACAGG 1373
Db 1490 TTTGGAGCAGCATCAATATCTGAGCTCCCGCTCCCGCTGCTGGGATCTATCCAAACCCAGT 1549
Qy 1374 CTGCGAGCTGCTATA-----TGCTTCCCATCTTGGACCTACTACTA 1417
Db 1550 GTCCCGGCTATCTTAACACCCACGCTGCTTACACTTTAGCGAGCTACTACTA 1602

RESULT 10
US-10-052-482-198
; Sequence 198, Application US/10052482
; Publication No. US200400722641
; GENERAL INFORMATION:

; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 198
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-052-482-198

Query Match 38.9%; Score 562.2; DB 12; Length 1359;
Best Local Similarity 66.4%; Pred. No. 2.9e-175;
Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;

Qy 63 ATGGCAAGCACTATTAAAGGAAGCATTTACAGTGGTGAAGAGACAGTCTTGTGTTGAG 122
Db 1 ATGGACGGGACATTTAAGGAGCTCTGTGCGTGGTGAAGGAGCAGGACGATCTTGTGAC 60
Qy 123 TGTGCTACGAGTCG---CCCCACCTTGCAGAGACAGAAATGACAGCTCTCTCTTCCAGT 179
Db 61 TCAGCGTACGAGCGCGCAGCCCATCTCCCAAGGCGACATGACTGCTTCGCGGAGTCTCT 120
Qy 180 GAATATGGGCAACATCAAGATGAGCGCGCTTCCCGACGAGACTGGTGTATCACAG 239
Db 121 GACTACGGGACGCCCAAGATCAACCCCTCCACACACAGGAGTGGATCAATCAG 180
Qy 240 CCCCCGGCCAGAGTTACCATTAAGATGGAGTGTAAACCCAAACAGGTTAATGGGTCAAG 299
Db 181 CC-----AGTGGGGTCAACGTCAAGCGGGAGTATGACCACATGAATGGATCCAG 231
Qy 300 AATTCACTCTGATGACTGCGAGCGTGGCAAGAGGAGGAAATGTTAGCAGTTTCAGACAAT 359
Db 232 GAGTCTCCCGTGGATGCGAGCGTTAGCAATGACGACAGCTGTGGCGGAGCGAGTCC 291
Qy 360 GTTGGATCAACTATGGAAGCTACATGGAAGAGAGCAT---ATTCGGCTCCAAATATG 416
Db 292 AACCCCATGACTACAAACAGTATATGGACGAGAAAGATGGCCCCCTCTCCCAACATG 351
Qy 417 ACACCAATGACGAGAGAGTATTGTGCCAGCAGATCTTACGTATGAGGACACAGACCAT 476
Db 352 ACCACCAACGAGGAGAGTCTGTCCCGCAGACCCCACTGTGGACACAGGAGCAT 411
Qy 477 GTACGGCAGTGGGTGGGAGTGAAGAGTATGCTTCCAGACGTGGACATCTTTG 536
Db 412 GTGAGCAATGGCTGGAGTGGCCATAAAGGAGTACAGCTTGATGGAGATCGACACATCC 471
Qy 537 TTGTTCCAGAACTTATGATGGGAAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGA 596
Db 472 TTTTTCAGAACTGATGGCAAGGAACTGTGTAAATGAACAGGAGGACTTCTCCCG 531
Qy 597 CTCAGCCGAGCTATAACGACAGATATCTCTGTTCACACTACACTACCTCAGAGAGAGA 656
Db 532 GCACCAACCTCTAACAACGAGAGTGTGTGTACACTCAGTTACCTCAGGGAAG- 590
Qy 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAAATAACA 716
Db 591 -----TTCACTGTGCTCTATAATAACAAGCTCCCAACCGACCAATCTCTCAGATTGAGT 645
Qy 717 ACAAGGCCAGTTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGAGAGTACAGCCAT 776
Db 646 GTCAAAAGAGACCTTCTTATGACTCAGTCAGAGGAGGAGCTTTGGGCAATACATGAT 705
Qy 777 CC---CACTCAGTCAAAAAGCTACCCAAACCATCATCTTCAACAGTGGCCCAAAACAGAGAC 833
Db 706 TCTGGCTCAACAAAGTCTCTCCCTTTGGAGGGGACAAACGATCAGTAAGAATACAGAG 765

```

QY 834 CAGCGTCTCAGTTAGATCCCTTATCAGATCTTGGACCGACAGAGCCGCTCTTGGCAAT 893
DB 766 CAACGGCCCCAGCCAGATCCGATATCAGATCTCGGGCCCCAGCCAGCAGTCCCTAGCCAAC 825
QY 894 CCAGGAGTGGGAGATACAGTATGCGAGTTCCTACTCTGAGCTTCTGTCGAGCAGCTCC 953
DB 826 CTTGGAGCGGGCAGATCCAGCTGTGGCAATTCCTCTGAGCTGCTCTCCGACAGGCC 885
QY 954 AACTCCAACTGATCAGCTGGGAGGGGCAAAATGGGAGGTTCAAGATGACAGACCTGAT 1013
DB 886 AACGCCAGCTGTATCACTCGGAGGGGACCAACGGGAGTTCAAAATGACGGACCCCGAT 945
QY 1014 GAAGTGGCTGGGCTTGGGAGAGAGGAAAGCAAACTTAACATGAATGAACAACTC 1073
DB 946 GAGTGGCCAGCCCTGGGGGACGCGAAGAGCCCAACATGAATTAAGACAGCTG 1005
QY 1074 AGCGTGCATCTCCTACTACTATGACAAAATATATGACTAAAGTTCAATGTAACGC 1133
DB 1006 AGCGGGCCCTCCGTTATTAATGATAAAACATTAATGACCAAGTGCACGGCAAAAGA 1065
QY 1134 TATGCTACAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGCCCTCAGCCCTCCAGRA 1193
DB 1066 TATGCTACAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGGCCCTCCAGCCCTCCAG 1125
QY 1194 TCATCCATGTACAAATACCCATCAGACCTCCCTTACATGAGTTCTTACCATGACACACCCC 1253
DB 1126 TCGTCCATGTACAAATACCCCTTCTGACATCTCTTACATGCTTCTTACCATGCCCCACG 1185
QY 1254 CAGAGATGAACTTTGAGTCCCATCCCTCCCTGCTTTGCGGTAAGCTCATCAGCTTT 1313
DB 1186 CAGAGGTTGAACTTTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1245
QY 1314 TTTGTGCCCCATATCCATCTAGGAATTCACCAACTGGAGGCACTTACCCCAATACCA-- 1371
DB 1246 TTTGGAGCGGATCACAATATCTGGACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1305
QY 1372 -----GGCTGCGAGCTGCTCATATGCTTCCCATCTTGGCACCCTACTACTA 1417
DB 1306 GTCCCCGCCATCTTAACACACCGCTGCTTCACTTACACTTAGCGACTACTACTA 1358

```

RESULT 11

```

US-10-052-482-197
; Sequence 197, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 197
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-197

```

Query Match 38.9%; Score 562.2; DB 12; Length 2957;
 Best Local Similarity 66.4%; Pred. No. 4.9e-175;
 Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;

```

QY 63 ATGGCAAGCAGCTATTAAAGAGCAATATCAGTGGTGAAGTGAAGCCAGTCTCTTTGAG 122
DB 173 ATGGACGGGAGCTATTAAAGAGGCTCTGTCGGTGGTGGTGGAGCGACGACCTCTCTTGAC 232

```

```

QY 123 TGTGCTACGGATCG---CCCCACCTTGCAGAGAGAGAAATGACAGACCTCTCTTCCAGT 179
DB 233 TCAGGTTACGGAGCGGAGCCCATCTCCCAAGGCGGACATGATGCTCTCGGGAGTCT 292
QY 180 GAATATGGGCAAAACATCAAGATGAGCCCGCGTTCGCCAGCAGGAGTGTGTATCAAG 239
DB 293 GACTACGGGCGAGCCCAACAGATCAACCCCTCCCAACACAGCAGGAGTGGATCAATCAG 352
QY 240 CCCCCGGCCAGAGTTACCATTAAGATGGAGTGAACCCAAACACAGGTTAATGGGTCAAGG 299
DB 353 CC-----AGTGAGGCTCAACGTCAGCGGTAGCAATGACAGAGCTGGTGGGCGAGGAGTCC 463
QY 300 AATTCACCTGATGACTGACGCGTGGCAAAAGAGGAGAAATGGTTAGCAGTTACAGCAAT 359
DB 404 GAGTCTCCGTTGAGCTGACGCGTTAGCAATGACAGAGCTGGTGGGCGAGGAGTCC 463
QY 360 GTTGGATGAATATGGAAGCTACATGGAAGAGAGCAT---ATTCCGCTCCCAATATG 416
DB 464 AATCCCATGAATCAACAGCTATATGACGAGAGAAATGGCCCCCTCTCCCAACATG 523
QY 417 ACAACCAATGAACGAAGAGTTATTGTGCGAGCAGATCTTACGTTATGGAGCAGACCAT 476
DB 524 ACCACCAAGAGAGAGTATCATGTCGCCGAGACCCACACATGTCGACACAGGAGCAT 583
QY 477 GTACGGAGTGGCTGGAGTGGGAGTGAAGAGTATGTTCTTCCAGAGCTGGACATCTTG 536
DB 584 GTGAGGCAATGCTGGAGTGGGCCATAAAGAGTACAGCTTGAAGAGATGACACATCC 643
QY 537 TTGTTCCAGAACATGATGGGAAGAGTTGTGTAATGAACAAAGATGACTTCCAGAGA 596
DB 644 TTTTTCAGAACATGATGGGAGAGTGTGTAATGAACAAAGAGTGTCTCTCCGCG 703
QY 597 CTCAGCGGAGTATAAGCAGATATCTCTCTGTCACACCTACACTACTCTCAGAGAGA 656
DB 704 GGCACACCTCTACACACAGAGTGTGTTGTCACACCTCAGTTACTCTCAGGAGAAAG- 762
QY 657 GGAGGCACCTTTATTTTCCAAATACATCAGTTTACCAGAGCAACGCAAGAGATAACA 716
DB 763 -----TTCACTGCTGGCTATATATACCTCCCAACCGCAATCTCTCAGATTGAT 817
QY 717 ACAAGCCAGATTTACCTTATGAGCAAGAGAGAGATCAGCGTGAAGAGTCAAGCCAT 776
DB 818 GTCAAGAGAGACCTTCTTATGACTCAGTCAAGAGAGAGCTTGGGGCAATTAACATGAT 877
QY 777 CG---CACTCAGTCAAAAGCTACCAACCATCATCTTCAACAGTGCCTCCCAACAGAGAC 833
DB 878 TGTGGCTCAACAAAGCTCTCTCTTGGAGGGGCAACACGATCAGTAAAGATACAGAG 937
QY 834 CAGCGTCTCAGTTAGATCTTTATCAGATTTTGGACCGGACAGCGCGCTTTGCAAT 893
DB 938 CAAGCGCCCCAGCCAGATCCGATCAGATCTCTGGGCGGACACAGCAGTCCGCTAGCCAC 997
QY 894 CCAGGAGTGGGAGATACAGCTATGCGAGTTCTTACTGGAGCTTCTGTCGACAGCTCC 953
DB 998 CTTGGAAGCGGGCAGATCCAGCTGTGGCAATTTCTCTGGAGCTGCTCTCCGACAGCGCC 1057
QY 954 AACTCCAACTGATCAGCTGGGAGGGGCAAAATGGGAGTTCAAGATGACAGACCTGAT 1013
DB 1058 AACGCGAGCTGTATCACTGGGAGGGGCAACAGGGGAGTTCAAATGACGACCCCGAT 1117
QY 1014 GAAGTGGCTGGGCTTGGGAGAGAGGAAAGCAAACTTAACATGAATGACAACTC 1073
DB 1118 GAGTGGCGAGCGCTGGGGCGAGGAAAGCAAGAGCCCAACATGAATTAACAGAGCTG 1177
QY 1074 AGCGGTGCACTTCGCTTACTACTATGACAAAAATATTTATGACTAAAAGTTTATGTTAAACCC 1133
DB 1178 AGCGGGCCCTCCGTTATTTACTATGATAAAACATTTATGACCAAGTGCACGGCAAAAGA 1237
QY 1134 TATGCTCAAAATTTGATTTCCAGGAATGCTCAGGCCCTCAGCCCTCAGCTCCAGAA 1193
DB 1238 TATGCTTACAAATTTGACTTCCAGGCAATTTCCAGGCTCTCGAGGCAATCCGACCGAG 1297
QY 1194 TCATCCATGTACAAATFACCCATCAGACCTCCCTCTACATGATGTTCTTACCATGACACCCC 1253

```

Db 1298 TCGTCAAGTAAAGTACCTTCTGACATCTCTACATGCTTCTTACCATGCCACAG 1357
 Qy 1254 CAGAAGATGAATTTGTAGCTCCCATCCCTCTCTTGTCCCGTAACCTCATCCAGCTTT 1313
 Db 1358 CAGAAGGTGAATTTGTCTCTCCCATCCATCTCTCCATGCTGCTCTCTCTCTCTCT 1417
 Qy 1314 TTTGCTGCTCTATTCATCTGAAATTCACCACTGGAGCATCTACCCCAATACCA-- 1371
 Db 1418 TTTGAGCGGCATCACAATCTGGACCTCCCTCCAGGGGGAATCTACCCCAACCCCAAC 1477
 Qy 1372 -----GGCTGCCAGCTGCTCATATGCTTCCCATCTTTGGACACTACTACTA 1417
 Db 1478 GTCCCCCGCATCTTAACACCCACGCTGCTTTCACATTTAGGACACTACTACTA 1530

RESULT 12
 US-10-342-887-710
 ; Sequence 710, Application US/10342887
 ; Publication No. US20040058340A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yudong
 ; APPLICANT: Linsley, Peter S.
 ; APPLICANT: Mac, Mao
 ; APPLICANT: Roberts, Christopher J.
 ; APPLICANT: Van 't Veer, Laura Johanna
 ; APPLICANT: Van de Vijver, Marc J.
 ; APPLICANT: Bernards, Rene
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE REFERENCE: 9301-188-999
 ; CURRENT APPLICATION NUMBER: US/10/342,887
 ; CURRENT FILING DATE: 2003-01-15
 ; PRIOR APPLICATION NUMBER: 60/298,918
 ; PRIOR FILING DATE: 2001-06-18
 ; PRIOR APPLICATION NUMBER: 60/380,710
 ; PRIOR FILING DATE: 2002-05-14
 ; PRIOR APPLICATION NUMBER: 10/172,118
 ; PRIOR FILING DATE: 2002-06-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 710
 ; LENGTH: 2957
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-342-887-710

Query Match 38.9%; Score 562.2; DB 13; Length 2957;
 Best Local Similarity 66.4%; Pred. No. 4.9e-175;
 Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;

Qy 63 ATGCAAGCACTATTAAAGAGCATATTCAGTGTGAGTGAAGACCACTCTTTTGTAG 122
 Db 173 ATGACGGGACTATTAAAGAGGCTCTGTCTGGTGTGAGCGACGACCACTCTTTGAC 232
 Qy 123 TGTGCTACGGATCG---CCCCACCTTGAAGACAGAAATGACAGCTCTCTTCCAGT 179
 Db 233 TCAGGTACGGAGCGGAGCCCATCTCCCAAGCGCGACATGATGCTCGGGAGTCT 292
 Qy 180 GAATATGGGCAACATCAAGATGAGCCCGCGCTTCCCGAGCAGGACTGGTTATCAG 239
 Db 293 GACTACGGGCGACCCCAAGATCAACCCCTCCCAACACAGCAGGAGTGGATCAATCAG 352
 Qy 240 CCCCAGGAGGTTACCATTAAGTGGAGTGTACCCCAACACAGGTTAATGGTCAAG 299
 Db 353 CC-----AGTGAGGCTCAACGCTCAAGCGGGAGTATGACCAATGAATGGATCCAGG 403
 Qy 300 AATTCACTGTATGACTGAGCGGTGGCAAAAGAGGGAATATGGTTAGCAGTTTCAGACAAT 359
 Db 404 GAGTCTCCGTGGAGTGTGACGGTTAGCAATATGACAGAGTGTGGGGGAGGCGAGTCC 463
 Qy 350 GTTGGGATGAATCTATGGAAGTGTATGGAAGAGAGCAT---ATTCCGCTTCAAAATG 416
 Db 454 AACCCCATGAATCAACAGCTATATGACGAGAGAGATGCGCCCTCTCTCCCAACATG 523

Qy 417 ACAACCAATGAACGAGATTTATTTGTCGAGAGATCTTACGTTATGGAGCAGACACCAT 476
 Db 524 ACCACCAACGAGAGAGAGTCACTGCTCCCGCAGACCCACACTGTGGACACAGGAGCAT 583
 Qy 477 GTACGGCACTGGCTGAGTGGGCACTGAAGAGTATGCTTCCAGACCTGGACATCTTTG 536
 Db 584 GTGAGCAATGCTGAGTGGGCACTAAGAGTACAGCTTGTGATGGAGATCGACATCC 643
 Qy 537 TTGTTCCAGAAATTTGATGGGAAAGTGTGTGTAATAAGCAAAAGATGACTTCCAGAGA 596
 Db 644 TTTTTCAGAAATGATGGAGAACTGTGTAAATGAACAAGAGGAGTCTCTCCG 703
 Qy 597 CTCAGCCGAGCTATAACGAGATATCTCTCTGTGTCACACTACACTTACCTCAGAGAGA 656
 Db 704 GCCACACCTCTTACACAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 762
 Qy 657 GGAGCCACTTTTATTTTCCAAATATCATCAGTTTACCCAGAGAGCAACGCAAAAGATAACA 716
 Db 763 -----TTCACTGTGGCTTATTAACAACCTCCACACCGCAATCTCTCAGATTGAGT 817
 Qy 717 ACAAGCCAGATTACCTTTATGAGCAAGAGAGATCAGCGTGGAGAGTCCAGGCCAT 776
 Db 818 GTCAAGAAGACCCCTTTTATGACTCAGTCAGAGAGAGGAGTCTGGGCAATTAATGAT 877
 Qy 777 CC---CACTCAGTCAAAAAGCTACCCAAACCATCATCTTCAACAGTGTCCCAAAAACAGAGAC 833
 Db 878 TCTGSCCTCAACAAAAGTCTCTCCCTTGGAGGGGCAAAACGATCAGTAAGAATAACAGAG 937
 Qy 834 CAGCGTCTCAGTTAGATCTCTTATCAGATTCTTTGACCGACGAGCAGCGCTCTTCAAT 893
 Db 938 CAACGCCCCCAGCAGATCCGTATCAGATCTCTGGGCGCCGACGAGCAGTCCGCTAGGCAAC 997
 Qy 894 CCAGGAGTGGGAGATACAGATACAGTATGGCAGTTCTCTACTGGAGTCTTGTCTGGACAGCTCC 953
 Db 998 CTTGGAAGCGGCGAGATCCAGCTGTGGCAATTTCTCTCTGGAGCTGTCTCTCCGACGCGC 1057
 Qy 954 AACTCCAAATGATCACTCTGGAGGGCAAAATGGGAGTTCAAGATGACAGACCTGAT 1013
 Db 1058 AAGCCAGCTGTATCACTCTGGAGGGGCAAAACGCGGAGTTCAAAATGACGAGACCCCGAT 1117
 Qy 1014 GAAGTGGCTCGGCTTGGGAGAGAGAAAGCAAAACCTAAACATGAACATATGACAAACTC 1073
 Db 1118 GAGTGGCCAGGCGCTGGGCGAGCGGAAAGCAAGCCCAACATGAATTCGACAGCTG 1177
 Qy 1074 AGCGTGCATCTGCTACTACTATGACAAAAATATATGACTAAAGTTTATGGTAAACGC 1133
 Db 1178 AGCGGGCCCTCGT-TATTACTATGATAAAACATTTATGACCAAGTGCACGGCAAAAGA 1237
 Qy 1134 TATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCGCTCCAGCTCCAGCTCCAGAA 1193
 Db 1238 TATGCTTCAAAATTTGATTTCCAGGCAATGCCAGCTCTCGAGCCATCGGACCGAG 1297
 Qy 1194 TCATCCATGTACAAATPACCCATCAGACCTCCCTACATGAGTTCTTACATGACACACCC 1253
 Db 1298 TGTGTCATGTACAAATPACCCATCAGACCTCTGACATCTCTCATGCTCTCTTACCATGCCAC 1357
 Qy 1254 CAGAAGATGAATTTGTAGTCTCCCATCCCTGTCTTTGCGCGTAACTTCTTCCAGCTTT 1313
 Db 1358 CAGAAGTGAATTTGTCTCTCCCATCCATCTCTCATGCTCTGCTCTCTCTCTCTCTCTCT 1417
 Qy 1314 TTTGCTGCCCCAATCCATCTGGAATTCACCAATGAGGAGCATCTTACCCCAATACCA-- 1371
 Db 1418 TTTGAGCGGCATCAATATCTGGACCTCCCAACGGGGGGAATCTTACCCCAACCCCAAC 1477
 Qy 1372 -----GGCTGCGAGCTGTCTATGATGCTTCCCATCTTTGGACACTTACTACTA 1417
 Db 1478 GTCCCCCGCATCTTAACACCCACGCTGCTTTCACACTTAGGAGCTTACTACTA 1530

RESULT 13
 US-10-172-118-710
 ; Sequence 710, Application US/10172118

```

RESULT 14
US-10-007-926A-295
; Sequence 295, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; TITLE OF INVENTION: USING ARRAYS OF CANDIDATE GENES
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 295
; LENGTH: 2957
; TYPE: DNA

```



```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18410
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000163.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUE 9.00e-99
; OTHER INFORMATION: EST HUMAN HIT: R87572.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: M17254.1, EVALUE 0.00e+00
US-09-864-761-18410

```

```

Query Match      26.7%; Score 387; DB 9; Length 567;
Best Local Similarity 83.7%; Pred. No. 2.9e-117;
Matches 438; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 896 AGGAGTGGGAGATACAGCTATGCGAGTTCCTACTGAGGCTTCTGCGAGACGCTCCAA 955
Db 523 AGGAGTGGGAGATCCAGCTTGGAGCTTCTCTCTGAGCTCTCTGCGAGACGCTCCAA 464

QY 956 CTCGAAGTGCATCACTGGGAGGCGACAAATGGGGAGTTCAAGATGACAGACCTGATGA 1015
Db 463 CTCGAAGTGCATCACTGGGAGGCGACCAACCGGGAGTTCAAGATGACAGACCTGCGAG 404

QY 1016 AGTGCTGGGCTTGGGAGAGAGAAAGCAACCTAACTGAACTATGACAAACTCAG 1075
Db 403 GGTGCGCGCGCTGGGAGAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 344

QY 1076 CGTGCACTTCGCTACTACTATGACAAATAATTATGACTAAAGTTATGTTAAACGCTA 1135
Db 343 CCGCGCCCTCGCTTACTACTATGACAAAGATCATGACCAAGGTCCATGGGAAGCGCTA 284

QY 1136 TGCCTACAAATTTGATTTCCAGGATCGCTCAGGCGCTCCAGGCTCAGCTCCAGATC 1195
Db 283 CGCTTACAAAGTTCGACTTCCAGGATCGCCAGGCTCCAGGCGCTCCAGGCTCCAGGCTC 224

QY 1196 ATCCATGTACAAATACCCATCAGACCTCCCTTACATGAGTTCTTACCATGACACCCCA 1255
Db 223 ATCTCTGTACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCTTACGCGCCACCCACA 164

QY 1256 GAAGATGAATTTGATGCTCCGATCCGCTCCGCTTTCGCGTACCTCATCGAGCTTTT 1315
Db 163 GAAGATGAATTTGTTGGGCGCCCGCCAGCTCCAGGCTCCCGGAGATCTTCCAGTTT 104

QY 1316 TGTGCCCCATATCCATACCTAGGAATTCACCAACTGGAGGATCTTACCCCAATACAGGCT 1375
Db 103 TGTGCCCCAAACCCATACCTAGGAATTCACCAACTGGGATATATACCCCAACACTAGGCT 44

QY 1376 GCCAGTGTCTATATGCTTCCCTTCCATCTTGGACCTACTACTAA 1418
Db 43 CCCACAGCCCATATGCTTCTCATCTGGGCACTTACTACTAA 1

```

Search completed: May 27, 2004, 21:11:29
Job time : 732 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2004, 16:28:06 ; Search time 4096 Seconds
(without alignments)
10549.457 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447

Sequence: 1 gaattccgcgaagcaataat.....gaaagaagcgccaagaaaa 1447

Scoring table: IDENTITY NUC
Gapop 10.0 ; Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_ges_hum.*

18: em_ges_inv.*

19: em_ges_pln.*

20: em_ges_vrt.*

21: em_ges_fun.*

22: em_ges_mam.*

23: em_ges_mus.*

24: em_ges_pro.*

25: em_ges_rod.*

26: em_ges_phg.*

27: em_ges_vrl.*

28: gb_gse1.*

29: gb_gse2.*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	917.2	63.4	1350	29	AY419454 Homo sapi
2	881.6	60.9	1866	11	AK050922 Mus muscu
3	879.4	60.8	1350	29	AY419456 Mus muscu
4	876.4	60.6	3064	11	AK076113 Mus muscu

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	856	59.2	1350	29	AY419455
6	692.4	47.9	726	13	BU334592
7	570.2	39.4	1706	11	AK042111
8	560.6	38.7	1347	29	AY418895
9	559	38.6	3118	11	AK036655
10	552.6	38.2	1347	29	AY418893
11	526.2	36.4	606	13	BU478556
12	501.2	34.6	1256	29	AY418894
13	485.4	33.5	980	13	BQ233264
14	483.4	33.4	709	14	CB265538
15	482.2	33.3	715	12	BG388025
16	475.4	32.9	642	12	B1713036
17	474.4	32.8	521	13	BU440626
18	468.6	32.4	701	9	AU136709
19	468.6	32.4	1089	13	BQ212127
20	462	31.9	634	12	BM489636
21	457.4	31.6	810	13	BU330912
22	452.2	31.3	668	10	BB680034
23	445.2	30.8	915	12	BG390291
24	444.2	30.7	928	12	BG259459
25	442	30.5	1045	12	BM456833
26	438.4	30.3	877	13	BU332566
27	437.4	30.2	567	13	BU336313
28	436.2	30.1	675	10	BB639043
29	432.6	29.9	629	10	BB577887
30	428.8	29.6	629	10	AM948986
31	425.6	29.4	592	12	BM088063
32	403.4	27.9	650	13	BU490135
33	400.2	27.7	616	14	CB579500
34	393.6	27.2	602	14	CB582161
35	389.8	26.9	2420	11	AK083900
36	385.8	26.7	526	14	CA405791
37	384.4	26.6	802	14	CB952600
38	380	26.3	827	9	AJ456498
39	367.8	25.4	522	10	BF566705
40	367.2	25.4	516	12	BG691566
41	361	24.9	693	13	BX100094
42	359.6	24.9	560	9	AU280022
43	358.8	24.8	976	13	BX328036
44	354	24.5	1406	28	CC264217
45	352.8	24.4	949	12	B1558012

ALIGNMENTS

RESULT 1	AY419454	1350 bp	DNA	linear	GSS 17-DEC-2003
LOCUS	AY419454				
DEFINITION	Homo sapiens ERG gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.				
ACCESSION	AY419454				
VERSION	AY419454.1				GI:39775411
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 1350)				
AUTHORS	Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Direct Submission				

JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA									
COMMENT	These sequences were made by sequencing genomic exons and ordering them based on alignment.									
FEATURES	Location/Qualifiers									
	1..1350									
gene	/organism="Homo sapiens"									
	/mol_type="genomic DNA"									
ORIGIN	/db_xref="taxon:9606"									
	<1..>1350									
Query Match	/gene="ERG"									
	/locus_tag="HCM6996"									
Best Local Similarity	63.4%; Score 917.2; DB 29; Length 1350;									
Matches 1668; Conservative	79.1%; 0; Mismatches 270; Indels 12; Gaps 2;									
QY	81	GAAGCATTATCAGTGGTGAAGACCAAGCCAGTCCCTGTTGTTGAGTGCGCTACGGAATCCGCC	140							
Db	1	GAAGCCTTATCAGTGTGAGTGAGGACCAAGTCTGTTGTTGAGTGCGCTACGGAATCCGCCA	60							
QY	141	CACCTTGCAAGACAGAAATGACAGCCCTCTTCCAGTGAATATGGCAACATCAAG	200							
Db	61	CACCTGGCTAAGACAGAGATACCGCTCTCTCCAGCAGCTATGAGACACTTCCAG	120							
QY	201	ATGAGCCGCGCTTCCCGCAGCAGACTGTTATCACAGCCCGCGCCAGAGTTACCAATT	260							
Db	121	ATGAGCCCGCAGCGTCCCTCAGCAGGATGGTGTCTCAACCCCGCAGCGGTCACCATC	180							
QY	261	AAGATGGAGTGAACCCAAAACAGGTTAATGGGTCAAGGAATTCACCTGATGATGCGAGC	320							
Db	181	AAATGGAAATGTAACTAGCCAGGAGTGAATGGCTCAAGGAACTCTCTGATGAATGCAAT	240							
QY	321	GTGGCAAAAGAGGAGAAATGGTTAGCAGTTTACAGCAATGTTGGGATGAATATGGAAGC	380							
Db	241	GTGGCCAAAGGCGGGAAGATGGTGGGAGCCAGACACCGTTGGATGAATCAAGCAGC	300							
QY	381	TATCGTGAAGAGACATATTCCTCGCTCCAAATATGACAACATGAAGAAAGAGTTATT	440							
Db	301	TATCGGAGGAGAGACATGTCACCCCGCAACATGACACGAAAGCGCAGAGTTATC	360							
QY	441	GTGCAGCAGATCTTACGTTATGAGAGACAGACCATGACGCGAGTGGCTGGAGTGGGCA	500							
Db	361	GTGCAGCAGATCTTACGTTATGAGTACAGACCATGTCGCGCAGTGGCTGGAGTGGGCG	420							
QY	501	GTGAAGGAGTATGGTCTTCCAGACGTGGACATCTTTGTTGTTCCAGAAATGATGGGAA	560							
Db	421	GTGAAGGAGTATGGCTTCCAGACGTCAACATCTTTGTTATTCAGAAATGATGGGAA	480							
QY	561	GAGTTGTGTAATGACCAAGATGACATTCAGAGACTCACCGAGCTATAACGCGAGAT	620							
Db	481	GAACTGTGCAAGATGACCAAGAGAGTTCAGAGCTCACCCAGCTACACGCGGAC	540							
QY	621	ATCTCTCTGTACACCTACCTACCTCAGAGAGA-----GAGGAGCCACTTTTATT	671							
Db	541	ATCTCTCTGTACACCTACCTACCTCAGAGAGANNNNNNNNNNNNNNNNNNNNNNNN	600							
QY	671	TTTCCAAATACATCAGTTTACCGAGACGACGCAAGAAATTAACACAGCCAGATTTA	731							
Db	601	NN	660							
QY	731	CCTTATGAGCAAGCGAGAGATCAGCGTGGAGAGTCAAGCCATCCCA---CTCAGTCA	788							
Db	661	CCATATGAGCCCGCCAGGAGATCAGCTGGACCGGTGACGCGCCACCCACGCCAGTCG	720							
QY	788	AAAGCTACCAACCATCATCTTCAACAGTGGCCCAAAACAGAGACCGGCTCTCAGTTA	848							
Db	721	AAAGCTGTCAACCATCTCTCTTCAAGTGGCCCAAAACAGAGACCGGCTCTCAGTTA	780							
QY	848	GATCTTATCAGATCTTGGACCGCAGCAGCGCTTGTGAAATCCAGGAGTGGGAG	908							
Db	781	GATCTTATCAGATCTTGGACCAACAGTACCGCTTGTGAAATCCAGGAGTGGGAG	840							

QY 909 ATACAGCTATGCGAGTCTCTACTGAGCTTCTGTGGACAGCTCCAACTCCAACTGCATC 968
Db 841 ATCCAGCTTTGGAGTCTCTCTGAGCTCTCTGTGGACAGCTCCAACTCCAACTGCATC 900
QY 969 ACCCTGGAGGACCAAAATGGGAGTTCAAGATGACAGACCCCTGATGAAGTGTCTGGCGT 1028
Db 901 ACCCTGGAGGACCAAAATGGGAGTTCAAGATGACAGACCCCTGATGAAGTGTCTGGCGC 960
QY 1029 TGGGAGAGAGGAAAGCAAACTAACATGAAGTATGACAACTCAGCGCTGCATCTGC 1088
Db 961 TGGGAGAGAGGAAAGCAAACTAACATGAAGTATGACAACTCAGCGCTGCATCTGC 1020
QY 1089 TACTACTATGACAAAATAATTATGACTAAAGTTTCAATGTTAAACCGTATGCTTACAAATTT 1148
Db 1021 TACTACTATGACAAAATAATTATGACTAAAGTTTCAATGTTAAACCGTATGCTTACAAATTT 1080
QY 1149 GATTTCCAGGAGTGTCTCAGCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTT 1208
Db 1081 GATTTCCAGGAGTGTCTCAGCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTT 1140
QY 1209 TACCCATCAGACTCTCCCTTACATGATGTTCTTACCATGTCACACCCCGCAGAGATGAATTT 1268
Db 1141 TACCCCTCAGACTCTCCCTTACATGATGTTCTTACCATGTCACACCCCGCAGAGATGAATTT 1200
QY 1269 GTAGCTTCCCATCTCCCTTGTGCGGTAACCTCATTCAGCTTCTTCTGCTGCTTCTTCTT 1328
Db 1201 GTGCGCGCCGCTCCCTCAGCTTCCAGCTTCCAGCTTCTTCTGCTGCTTCTTCTTCTT 1260
QY 1329 CCATCTGGAATTCACCAACTGAGGAGTCTTACCCCAATACCAAGTGCAGTGTCTCAT 1388
Db 1261 CCATCTGGAATTCACCAACTGAGGAGTCTTACCCCAATACCAAGTGCAGTGTCTCAT 1320
QY 1389 ATGCTTCCCATCTTGTGCGACCTTACTACTAA 1418
Db 1321 ATGCTTCTCATCTGCGCACTTACTACTAA 1350

RESULT 2
LOCUS AK050922 1866 bp mRNA linear HTC 20-SBP-2003
DEFINITION Mus musculus 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone:D030036124 product:Erp mRNA, mouse homolog of Human ets-related gene ERG, transcript variant 1, full insert sequence.
ACCESSION AK050922
VERSION AK050922.1 GI:26341597
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636
REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20493374
PUBMED 11042159
REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 MEDLINE
 PUBMED
 11076861
 4
 REFERENCES
 AUTHORS
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 REFERENCES
 AUTHORS
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 1866)
 REFERENCES
 AUTHORS
 Adachi, J., Aizawa, K., Akimura, T., Atakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toyota, I., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://phantom.gsc.riken.go.jp/
 Location/Qualifiers
 1..1866
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:D030036124"
 /db_xref="MGI:2418825"
 /db_xref="taxon:10090"
 /clone="D030036124"
 /tissue_type="whole body"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="9 days embryo"
 249..1637
 /note="unnamed protein product; Erg mRNA, mouse homolog of
 Human ets-related gene ERG, transcript variant 1
 (DBJAB073078, evidence: BLASTN, 99%, match=1768)
 putative"
 /codon_start=1
 /protein_id="BAC34461.1"
 /db_xref="GI:26341598"
 /translation="MIQVDPAAHIKEALSIVSDQSLFECAYGTPLAKTEMTASS
 SSDYGTSMKPRVQQDLWLSOAPARVTKMECPNSQVNGSRNPDSCVNVKGMKMG
 SDPTVGMVSGVSWKXVPPNNMTNERRVLPADPTLWSTDHVRLEWAKVYGLL
 DVDFLQNDGKLECKTKQDFQRLTFSYNADILLSHLYLRSTPLPRLTSDVDKA
 LQNSRLMHRANTDLPYEPFRKSAWGHSLTTPQSKAAQPSFSAVPKEDQRFQDPY
 QILGPTSRSLNPGSGQIQWQFLELLSDNSNCITWETNGEFGKMDDEVDARW
 GERKSKPNMNYDKLSRLRYDYDKNIMTKVHKRYAYKFPDFHGAQALQPHPPSSSLY

Query Match 60.9%; Score 881.6; DB 11; Length 1866;
 Best Local Similarity 79.3%; Pred. No. 2e-252;
 Matches 1075; Conservative 0; Mismatches 269; Indels 12; Gaps 2;
 ORIGIN
 QY 74 TATTAAAGGAAGATTATCAGTGGTGAAGAACACCAAGTCTCTTGTGTTGAGTGTGCTTACCG 133
 Db 281 TATTAAAGGAGGCTTGTCTAGTTGTGAGCGAGGACCACTATTTGAGTGTGCTTACCG 340
 QY 134 ATCGCCCAACCTTGCAAGACAGAAATGACAGCTCTCTTCCAGTGAATATATGGGCAAAAC 193
 Db 341 AACCCACACCTGGCTTAAGACAGAGATGACCGCATCTCTTCCAGTGAATATATGGCAGAC 400
 QY 194 ATCAAAAGATGAGCCCGCGTTCCTCCAGCAGGACTGGTATTCACAGCCCCCGGCGAGAGT 253
 Db 401 ATCAAGATGAGTCCCGAGAGTCCCTCAGCAGGACTGGTGTCTCAAGCCCCGAGGAGGT 460
 QY 254 TACCATTAAAGTGGAGTGTAAACCCAAACCAAGCTTAATGGTCAAGGAATTCACCTCATGA 313
 Db 461 CACCATCAAGATGGAGTGTCAACCCCTAGTCAGGTGAATGGTTCCAGGAATTCACCTCATGA 520
 QY 314 CTGCAGCGTGGCAAAAGAGGAGGAAAATGGTTAGCAGTTTCCAGACAAATGTTGGGATGAAC 373
 Db 521 GTGCAGTGTGAACAAAGGTGGGAAGATGTTGGSCAGCCCGGATCTGTGGGGATGAGCTA 580
 QY 374 TGGAACTACATGGAAGACAGAGCATATTCGGCTCCAAATATGACACCAAGTGAACGAAG 433
 Db 581 CGGCAGCTACATGGAGGAGAGCATGTGCGCGCTCCCAATATGACCAAAATGAGCGCAG 640
 QY 434 AGTTATTGTGCCAGCAGATCTCTACGTTATGGAGCAGACACCATGTACGGCAGTGGTGA 493
 Db 641 AGTGATCGTCCCTGCAGATCTCTCTGTGGAGCAGACCATGTCCGACAGTGGCTGGA 700
 QY 494 GTGGGAGTGAAGGAGTATGGTTCCACAGCTGGACATCTTGTGTTCCAGAACATTGA 553
 Db 701 GTGGGCGGTGAAGAATATGGCCCTCTCGATGTGGACGCTTACTATTTTCAGAAATATCA 760
 QY 554 TGGGAAGAGTGTGTGTAATAATGACAAAGATGACTTCCAGAGACTCACGCGGAGCTATAA 613
 Db 761 TGGGAAGAGCTGTGCAAGATGACAAAGATGACTTCCAGCGGCTCACGCGGAGCTACAA 820
 QY 614 CGCAGATATCTCTCTGTCAACCTTACATCTCTCAGAGAGACTCCCTTCCATCATCTGAC 664
 Db 821 TGCCGACATTTCTCTCTCATCTCCATCTCCACTACCTCAGAGAGACTCCCTTCCATCATCTGAC 880
 QY 665 TTTTATTTTCCAAATACATCAGTTTACCAGAGCAAGCGCAAGAAATACCAACAAAGGCC 724
 Db 881 TTCGATGAGCTGTGATAAGGCTTTACAAACTCTCCACGGTTAATGATGCCAGAAACAC 940
 QY 725 AGATTTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTCAAGCCATC---CCAC 781
 Db 941 AGATTTTACCTTATGAGCCTCCAGAGATCAGCGTGGACCGGCGACAGCCACCTCACC 1000
 QY 782 TCAGTCAAAAGCTACCCCAACCATCTTCAACAGTGGCCCAAAACAGAGACAGCAGGCTCC 841
 Db 1001 TCAGTCAAAAGCTGCTCAGCCATCTCCCTCTGAGTGGCCCAAAACTGAGACAGCGTCC 1060
 QY 842 TCAGTTAGATCTTATCAGATTCTTGGACCGACAGCAGCGCTTTCGAAATCCAGGAG 901
 Db 1061 TCAGTTAGATCTTATCCAGATCTTGGACCGACAGCAGTAGCGCGCTTGTCTAATCCAGGTAG 1120
 QY 902 TGGCAGATACAGCTATGGCAGTCTCTTCTGAGCTTCTGCGGACAGCTCCCACTCCAA 961
 Db 1121 TGGCCAGATCCAGCTGTGGCAGTCTCTGCTCGCAACTCTCTGACAGAGCTCCCACTCAA 1180
 QY 962 CTGCATCACCTGGAGGGGCAAAATGGGAGTTCAGAGTACAGACAGCCCTCATGAAGTGGC 1021
 Db 1181 CTGCATCACCTGGAGGGCACCACCGGGAGTTCAGATGACAGACCGCGAGCGAGTGGC 1240
 QY 1022 TCGCGGTTGGGAGAGAGGAAAAAGCAACCTAATCATGAACATATGACAACTAGCCGCTGC 1081

[illegible]

AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitho,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/ Location/Qualifiers 1. .3064 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:6330501C04" /db_xref="MGI:239312" /db_xref="taxon:10090" /clone="6330501C04" /sex="male" /tissue_type="medulla oblongata" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 61. .1431 /note="unnamed protein product; Erg mRNA, mouse homolog of Human <i>esr</i> -related gene <i>ERG</i> , transcript variant 1 (DBJ AB073078, evidence: BLASTN, 99%, match=1768) putative" /codon_start=1 /protein_id="BAC37131.1" /db_xref="GI:26346965" /translation="WASPIKEALSVSRDQLPECAYGTPLAKTEMTASSSSDYGT SYMSRVFQQHVPSPARVITVFPADPTLMSTDHVROWLEWAWKEYGLLDVDFLL KIDGCEKTKTDQFRLTSYNADILLSHLHIRELPHTLSDVDKALQNSRL MHARTGGAAFTFPNTYSPATQITTRPAAQSPSPAVPTEQDQRPDPQVILGET SNRLANPGQIGLWQFLELLSDSGNCTWIEGTGFKMTDPDEVARWGERKSK PNMVNDKLSRLRYVDKNTMKVHGKYAVYDFHGI AALQPHPPSSLYKYPDDL PYMGSYHAHPQMWNVSPHPALPTSSFFPSPNPYNSPTGGIYENTRLPASHMES HLGTYYI"
FEATURES	source
ORIGIN	Query Match 60.6%; Score 876.4; DB 11; Length 3064; Best Local Similarity 77.1%; Pred. No. 9.1e-251; Matches 1101; Conservative 0; Mismatches 306; Indels 21; Gaps 2; QY 9 GCGAGCAATAATTATTATTCAGCAATTATTAGCGATCAATAATCTTGATCAGATTATGCC 67 Db 6 GAGAAGCAATAATTATTATTAACATTATTAAACGATCAATAAATCTGATTGCATTATGCC 65 QY 68 AAGCACTATTAAAGGAGCAATTATCAGTGTGTGAGTGAGGACCAGTCCCTTGTTTCAGTGTGC 127 Db 66 CAGCACTATTAAAGAGGCCCTTGTGAGTGTGAGCGAGGACCAGTCACTATTGAGTGTGC 125 QY 128 CTACGAGATCGCCCACTTTCGAAAAGACAGAAATGACAGCCCTCCTCTTCCAGTGAATAATGG 187

126 CTACGGACGACACCTGGCTAAGACAGAGATGACCGCATCTCTTCCAGTACTATGG 185
188 CCAACATCAAGATGAGCCGCGTTCGCCAGCAGGACTGTTATACAGAGCCCGGC 247
186 CCAGACATCCAAAGATGAGTCCAGAGTCCCTCAGCAGGACTGGTGTCTCAAGCCCGC 245
248 CAGAGTTACCATTAAGATGGAGTGAACCCAAACCGAGTTAATGGGTCAAGGATTCAC 307
246 CAGGCTACCATCAAGATGGAGTGAACCCCTAGTCAAGTGAATGGTTCAGGAACTCAC 305
308 TGATGACTGCGAGCGTGGCAAAAGAGGAGAAAGTGTAGCAGTTTCAGACAAATGTGGAT 367
306 TGATGAGTGCAGTGTGAACAAAGTGGAGATGGTGGCAGCCCGGATACTGTGGGGAT 365
368 GAATCTAGAGCTACATGGAGAGAGCATATTTCCGCTCCCAATATGACACCAATGA 427
366 GAGTACGGCAGCTACATGGAGGAGAGCATGTCCGCTCCCAATATGACCAATGA 425
428 ACGAAGAGTTATTGTCCAGCAGATCTTACGTTATGGAGCAGACAGACCATGTAGCGCAGTG 487
426 GCGCAGAGTATCGTCCCTGAGATCTTACTCTGTGGAGCAGACCATGTCCGACAGTG 485
488 GCTGAGTGGCAGTGAAGAGTATGTTTCCAGAGCTGAGCAATCTTGTGTTCAGAA 547
486 GCTGAGTGGCGGTGAAGAAATATGGCTCTCTCGATGTGACGTCTTACTATTTCAGAA 545
548 CATTGATGGAAAGAGTTGTCTAAATGACCAAGATGACTTCCAGAGACTTCAGCCGAG 607
546 TATCATGAGGAGAGCAGTCAAGTCAAGTCAAGAGATGACTTCCAGCGGTTCAGCCGAG 605
608 CTATAACCGCATATCTCTCTGTCTCACACTACATCTACGAGAGAGAGAGCCACTTT 667
606 CTACATCCGACATCTCTCTCACATCTCCATCTACGAGAGAGACTCCCTTCCACA 665
668 TATTTTCCAAATACATCAGTTTACCAGAGCAA-----CGCAAG 709
666 TCTGACTTCCGATGACGTTGATAGGCTTTTCAAACTCTCCAGGTTAATGATGCCAG 725
710 AATACAAACAAGGCGAGATTTACCTTATGACCAAGCAGGAGATCAGCGTGGACGATCA 769
726 AAACACAGGGGTGCAGCTTTTATTTTCCAAATACTTCAGTATATCCGAAGTACGCA 785
770 CAGCCATCCACTCAGTCAAAAGTACCCACCATCATCTTCAACAGTGCCTCAACAGAA 829
786 AAGAAATCAACTAGGCC--AGTGTCTAGCCATCTCCCTCTCAGTGGCCCAAACTGA 842
830 AGACAGAGTCTCTCAGTTAGATCTTATCAGATTCTTGGACCGACAGCAGCGCTTTC 889
843 AGACAGAGTCTCTCAGTTAGATCTTACAGATCTCTGGACCGACAGTAGCGCTTTC 902
890 AATCCAGGAGTGGGAGATACAGTATGCGAGTTCTTACTGAGAGTTCTGTGGAGCAG 949
903 TAATCCAGGTATGGCCAGATCCAGCTGTGGCAGTTCTGTCTGAATCTCTGTGACAG 962
950 CTCCAACTCCAACTGCATCAGCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGACCC 1009
963 CTCCAACTCCAACTGCATCAGCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGACCC 1022
1010 TGATGAATGGCTCGCGCTTGGGAGAGAGAGAAAGCAAACTTAAGTAACTATGACAA 1069
1023 GGACAGGTGGTCTCGCGCTGGGAGGAGGAGCAAGCAAGCCCAATGAATGATGACAA 1082
1070 ACTCAGCGCTGACATCTCTACTACTATGACAAATAATATGACTAAAGTTCAATGTA 1129
1083 GCTCAGCGCGCTCTCCCTACTACTACGACAAATAATATGACTAAAGTTCAATGTA 1142
1130 AGCTATGCTCAAAATTTGATTTCCAGGAATCGCTCAGCGCTTCCAGCTTCACTCC 1189
1143 GCGCTACGCTTACAGTTTGTACTTCCAGGATTTGCCAGCGCTTCCAGCGCTTCCAG 1202
1190 AGAATATCCATGACAAATACCATCCATCAGACTCCCTTACATGATGTTCTTACATG 1249

1203 TGAGTCTGCTCCCTGTACAAAGTACCCCTCCGAGCTGCCATACATGGGCTCTATACGCCCA 1262
1250 CCCCAGAGATGAATTTGTAGTCTCCCAATCCCAATCCCTTGTGCTTCCGTAACCTATCCAG 1309
1263 CCCCAGAGATGAATTTGTGCTCTCCCAATCCCAATCCCTTGTGCTTCCGTAACCTATCCAG 1322
1310 CTTTTTGTGCTCCCTTAATCCATCTCATATGGAATTCACAACTGGAGCATCTACCCCAATAC 1369
1323 TTTCTTGTCTCCCGAACCACATCTGGAATTCACCGACTGGGCACTTACCGCAACAC 1382
1370 CAGGTGCGAGCTGTCTATATGCTTCCCAATCTTGGCACTTACTACTA 1417
1383 TAGGCTCCAGCCCATATGCTCTCTCAGCTGGCACCTACTACTA 1430

RESULT 5
AY419455 1350 bp DNA linear GSS 17-DEC-2003
LOCUS Pan troglodytes ERG gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY419455
VERSION AY419455.1 GI:39775412
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
AUTHORS Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1350)
AUTHORS Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1350
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
41..>1350
/gene="ERG"
/locus_tag="HCM6896"
ORIGIN
Query Match 59.2%; Score 856; DB 29; Length 1350;
Best Local Similarity 74.0%; Pred. No. 8e-245;
Matches 999; Conservative 0; Mismatches 339; Indels 12; Gaps 2;
QY 81 GAAGCATATCATGTGTGAGTGAAGACCAAGTCTCTTGTGAGTGTGCTACGGATCGCC 140
DB 1 GAAGCCTTATCATGTGTGAGCGAGGACCAAGTCTCTTGTGAGTGTGCTACGGACGCCA 60
QY 141 CACCTTGAACACAGAAATGACAGCTCTCTTCCAGTGAATATGGGCAACATCAAG 200
DB 61 CACCTGGCTTAAGACAGAAATGACAGCTCTCTTCCAGCACTATGGACAGACTTCCAG 120
QY 201 ATGAGCCCGCGTTCGCCAGCAGAGTGGTTATCACAGCCCCCGGCGCAGATTACCT 260
DB 121 ATGAGCCCAAGCGTCCCTCAGCAGAGTGGTGTCTCAACCCCGCAGCCAGGCTCACC 180
QY 261 AAGATGAGTGTATACCCAAACCAAGTGAATGGTCAAGGAATTCACCTGACTGCAGC 320


```
QY 826 CAGAAGACAGCGTCTCAGTTAGATCCTTATCAGATTCCTTGAGCCGACGACGCGTC 885
Db 129 CAGAAGACAGCGTCTCAGTTAGATCCTTATCAGATTCCTTGAGCCGACGACGCGTC 188
QY 886 TTGCAATCCAGGAGTGGGAGATGACGATGATGGAGTTCCTACTGAGCTTCTGTCGG 945
Db 189 TTGCAATCCAGGAGTGGGAGATGACGATGATGGAGTTCCTACTGAGCTTCTGTCGG 248
QY 946 ACAGCTCCAACTCCAACTGCATCCCTGGGAGGCGCAAAATGGGGAGTTCAAGATGACAG 1005
Db 249 ACAGCTCCAACTCCAACTGCATCCCTGGGAGGCGCAAAATGGGGAGTTCAAGATGACAG 308
QY 1006 ACCCTGATGAAGTGGCTCGGCTGGGAGAGAGAGAAAGCAAACTTAAGTAACTATG 1065
Db 309 ACCCTGATGAAGTGGCTCGGCTGGGAGAGAGAGAAAGCAAACTTAAGTAACTATG 368
QY 1066 ACAAACTCAGCGTGCATTCGCTACTACTATGACAAAATATTATGACTAAATGTTTCATG 1125
Db 369 ACAAACTCAGCGTGCATTCGCTACTACTATGACAAAATATTATGACTAAATGTTTCATG 428
QY 1126 GFAAAGCTATGCTTACAAATTTGATTTCCAGGAATCGCTCAGGCGCTCCAGGCTCACC 1185
Db 429 GFAAAGCTATGCTTACAAATTTGATTTCCAGGAATCGCTCAGGCGCTCCAGGCTCACC 488
QY 1186 CTCAGATCATCCATGACAAATACCATCAGACCTCCCTTACATGATGATTCCTTACCATG 1245
Db 489 CTCAGATCATCCATGACAAATACCATCAGACCTCCCTTACATGATGATTCCTTACCATG 548
QY 1246 CACACCCCGAGAGATGAACCTTTGAGTCCCATCCCTGCTTCCCGGTAACTCAT 1305
Db 549 CACACCCCGAGAGATGAACCTTTGAGTCCCATCCCTGCTTCCCGGTAACTCAT 608
QY 1306 CCAGCTTTTGTGCTCCCTTAATCATCTCGAATTCACCAACT-GGAGGCATCTACCCC 1364
Db 609 CCAGCTTTTGTGCTCCCTTAATCATCTCGAATTCACCAACTGGAGGCATCTACCCC 669
QY 1365 AATACGAGCTGCCAGCTGCTCATATGCTTCCCATC-TTGGCAGCTACTACTAAGTG 1421
Db 669 AATACGAGCTGCCAGCTGCTCATATGCTTCCCATC-TTGGCAGCTACTACTAAGTG 726

RESULT 7
AK042111
LOCUS
DEFINITION
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
library, clone:A630059111 product:Friend leukemia integration 1,
full insert sequence.
AK042111
VERSION
AK042111.1 GI:26334936
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
Carninci, P. and Hayashizaki, Y.
1
AUTHORS
High-efficiency full-length cDNA cloning
2
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
REFERENCE
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
2
AUTHORS
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
3
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159
REFERENCE
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
3
AUTHORS
Konno, H., Akiyama, J., Nishi, K., Katsurai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
```

```
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1706)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Sainsgawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC Building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
FEATURES
Location/Qualifiers
1..1706
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A630059111"
/db_xref="MGI:2407136"
/db_xref="taxon:10090"
/clone="A630059111"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
158..1516
/notes="unnamed protein product; Friend leukemia
integration 1 (MGD|MGI:95554, GB|NM_008026, evidence:
BLASTN, 99%, match=1671)
putative"
/codon_start=1
/protein_id="BAC31169.1"
/db_xref="GI:26334937"
CDS
```

/translation="MDGTIKELAVSVDDSDSLFDSAYGAAHLPKADMTASGSPDYGO
PHKINLPPOQEMINQPVNVNREYDHMSRESVDCSVSKNKLKVGGEANPMNY
NSYMDKXGPRPPNMTNERRVUPADPTLTQSHVROKLEWAKIKEYGLMEIDTSFO
NMCKELCKANKEDEFLAATSAYNTEVLLSHLSYLRSSLLAYNTTSHHTQSSRLNVE
DPSVDSVRGAANNMNSGLNKPLIGSSQTMGNTEQRPDPYQILGPTSSRLANP
GSGOIQWQFLLELLSDSANASCTITWETGTFBFTDDEVARWGERKSPNNDYK
LSRALRYFYDKNIMTKVGRYAYKFDHGIQAOLQPHPTTSMYKPSDISYMPYSYH
AHQOKNVFVPSHPSSMBVTSSPFGAASQYWTSPTAGIYPNPSPRHPNTHVPSHLGS
YY"

ORIGIN

Query Match	39.4%; Score 570.2; DB 11; Length 1706;
Best Local Similarity	66.8%; Pred. No. 5e-159;
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;	
QY	63 ATGCGACAGCATTATTAAGGAGCATTATCAGTGTGTGAGTGAAGACCACTGCTTGTGTTGAG 122
DB	158 ATGACGGGACTATTAGGAGCTCTGTCTGTGTGTGAGTGCAGATCACTCCCTTTTGTAT 217
QY	123 TGTGCTTACCGATCG----CCCCCTTGCAGAAAGACAGAAATGACAGCCTCTCTTTCAGT 179
DB	218 TCAGCATACGGAGCGGAGCCCATCTCCCAAGGCAGATATGATGCTTTCGGGAGTCCCT 277
QY	180 GAATATGGCAAAACATCAAGATGAGCCCGCGCTTCCCAAGCAGGACTGTTATCAGAG 239
DB	278 GACTACGGGAGCCCAACAAATCAACCCCTGCGCACCGCAGAGGAGTGGATCAACCGAG 337
QY	240 CCCCAGCCAGAGTTACCATTAAGATGGAGTGTAAACCCAAACCCAGGTTAATGGGTCAAGG 299
DB	338 CC-----AGTGAGAGTCAATGTCAAGCGGAGTATGACACATGAATGGATCCAGG 388
QY	300 AATTCACTGTATGACTGAGCGGTGGCAAGGAGGGAATGTTAGCAGTTCAGACAT 359
DB	389 GAGTCTCCGGTGGAGTGCAGTGTCAAGAAATGTAAACAGCTGTGGCGGAGGGAAGCC 448
QY	360 GTTGGGATGAATATGAAGCTACATGAAGAGAGCA--TATTCCGCTTCCAAATATG 416
DB	449 AACCCCATGAATATATAGTCTATGATGAGAGAGACGCGCCCTCTCTCCCAACATG 508
QY	417 ACAACCAATGAACGAGGTATTGTGCGCAGCATCTTACGTTATGAGGACAGACCAT 476
DB	509 ACCACCAACGAAAGAGAGTCAATTGTGCTTGCAGACCCCACTGTGGACACAGGAGCAC 568
QY	477 GTACGGAGTGGTGGAGTGGGAGTGAAGAGTATGCTTTCAGACGCTGGACATCTTG 536
DB	569 GTTCAAGTGGTGGAGTGGGCTATTAAGAAATACGGAATGATGGAGATGACACTTCC 628
QY	537 TTGTTCCAGAACATGTATGGAAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGA 596
DB	629 TTTCTTCCAGAACATGTATGGTGAAGGAAATGTGTAAATGAACAGGAGGAGTCTCTCCGA 688
QY	597 CTCAGCGGAGCTATAACGAGATATCTCTGTCACACCTACACTACCTCAGAGAGAGA 656
DB	689 GCCACCTCCGCTCAACACAGAGTGTCTGTGCGACCTCAGTTTACCTCAGGGAAG- 747
QY	657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAAATACA 716
DB	748 -----TTCACCTGCTGGCCTATAACACAACTCCCATACAGACCACTGCTCTCAGCACTGAAT 802
QY	717 ACAAGCCAGATTTACCTTATGACCAAGCGAGGAGTACGCGTGGAGGAGTCAACA--GC 773
DB	803 GTCAAGGAGAGCCCTTCTTATGACTCTCTCAGGAGAGGAGCATGGAACAATATATGAAC 862
QY	774 CATCCCACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGGCCCAACAGAGAGAC 833
DB	863 TCTGGCCTCAACAAAGCTCTCTCTCTGGAGATCACAGACCATGGGCAAGAACACTGAG 922
QY	834 CAGCGTCTCAGTTAGATCTTATCAGATTTCTTGGACCGACGAGCGGCTTTCGAAT 893
DB	923 CAGCGGCCCAAGCCAGATCTTATCAGATCTCTGGGGCCCAACAGCAGCGCGCTAGCAAC 982
QY	894 CCAGGAGTGGGACATACAGTATGGCAGTCTCTTACTTGGAGCTTCTGTCCGACAGCTCC 953

DB	983 CTTGGAGTGGGCGAGATCAGCTGTGGCAGTTTCTCTCGGAACACTGTCTCGACAGCGCC 1042
QY	954 AACTCCAACTGCATCACTCGGAGGGGCAAAATGGGGAGTTTCAAGATGACAGACCTGAT 1013
DB	1043 AACGCCAGCTGTATCACTTGGAGGGGACCAACGGGGAGTTTCAAAATGACGAGCCCTGAT 1102
QY	1014 GAAGTGGCTTGGCGGTGTGGGGAGAGAGAAAAGCAAACTTAACATCAATGAACTATGACAAATC 1073
DB	1103 GAGTGGCCAGCGCTGTGGGAGAGCGGAAGAGCAAGCCCAACATGAATATGACAGCTG 1162
QY	1074 AGCGTGACATTTGCTGCTACTTATGACAAAATATATGACTTAAAGTTTCAATGGTAAACGC 1133
DB	1163 AGCCGGGCGCTCCGATACTACTATGACAAAACATATTAAGCAAAAGTGTGTCGAAAGG 1222
QY	1134 TATGCTTACAAATTTGATTTCCACGGAATTCGTCAGGGCCCTCCAGCCTCACCTCCAGAA 1193
DB	1223 TATGCTTACAAATTTGATTTCCATGCAATTTGCCAGGCGCTCGACCCACATCCAAAGAG 1282
QY	1194 TCATCCATGTACAAATACCATCCATCAGACCTCCCTACATGAGTTCTCCATCAGCAGCC 1253
DB	1283 ACATCCATGTACAAATATCCCTCTGATATCTCTACATGCTTCTTACCTGCCATCAA 1342
QY	1254 CAGAAGATCAAACTTTGTAGTCTCCCATCCCTGCTTTTGGCCGTAACCTCATCCAGTTT 1313
DB	1343 CAGAAGTCAAACTTTGTCCGCTCTCAGCCATCTCCATGCTGTCACTCTCCAGCTTC 1402
QY	1314 TTTGCTGCCCTTAATCCATCTGGAATTCACCACTGAGGAGCTTACCCCATACCAGG 1373
DB	1403 TTTGGAGCAGCATCAATCTACTGAGCTCCCTCCCATCTGTGGGATCTATCCAAACCCAGT 1462
QY	1374 CTGCGAGCTGCTCATA-----TGCCTTCCCATCTTTGGCAGCTTACTACTA 1417
DB	1463 GTCCCCCGCATCTTAACACCCACAGTGTCTTACACTTAGGAGCTACTACTA 1515

RESULT 8
AY418895 1347 bp DNA linear GSS 17-DEC-2003
LOCUS Mus musculus FLII gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY418895
VERSION AY418895.1 GI:39774855
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Fierli, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Fierli, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source 1. .1347
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
gene <1. .>1347
/gene="FLII"

```
ORIGIN
Query Match 38.7%; Score 560.6; DB 29; Length 1347;
Best Local Similarity 66.7%; Pred. NO. 3.4e-156;
Matches 905; Conservative 0; Mismatches 419; Indels 33; Gaps 6;

QY 79 AGGAAGCATTTATCAGTGTGAGTGAAGACCAAGTCCCTGTTGTTGAGTGTGCTTACGGATCGC 138
DB 5 AGGAGGCTCTGTCTGTGTGTGAGTGAAGATCAGTCCCTTTTGTATCAGCATACGGAGCGG 64
QY 139 ---CCACCTTGCAGAAAGACAGAAATGACAGCCTCTCTTCCAGTGAATATGAGGCAACAT 195
DB 65 CAGGCCATCTCCCAAGGAGATATGACTGCTTCGGGAGTCTCTGACTAGCGGAGGCC 124
QY 196 CAAGATCAGCCGCGCTTCCACAGAGGACTGTTATACAGCCCGCCGAGAGTTA 255
DB 125 ACAAAATCAACCCCTGCGACGAGAGAGTGAATCAACAGCC-----AGTGA 175
QY 256 CAATTAAGATGAGTGTAAACCAACAGGTTAATGGGTCAAGAAATCACTGATGACT 315
DB 176 GAGTCAATGTCAAGCGGAGTATGACCACATGAATGGATCCAGGAGTCTCCGGTGGACT 235
QY 316 CGAGCGTGGCAAGAGGAGGAAATGTTAGCAGTTTACAGATTCAGACAATGTTGGGATGAATG 375
DB 236 CGAGTGTGAGCAATGTAAAGCTGTGTGGGAGGAGCGGAGCCAAACCCATGAATATA 295
QY 376 GAAGCTACATGAGAGAGAGCA---TATTCGCTCCCTCCAAATATGACAAACCAATGAACGA 432
DB 296 ATAGCTACATGATGAGAAGAGCGGCCCTCTCTCCCAATGACACCAACCAACGGA 355
QY 433 GAGTTATTGTGCCACAGATCCTAGTTATGGAGCAGACATGATGAGGAGTGGGTGG 492
DB 356 GAGTCATTGTGCTCGACACCCACACTGTGTGACACAGAGGACGTTGACAGTGGGTGG 415
QY 493 AGTGGGAGTGAAGGAGTATGTTCTTCAGAGCTGGACATCTGTTGTTCCAGAACATTTG 552
DB 416 AGTGGGCTATAAGGAATACGGAATGATGGAGATTGACACTCTCTTCCAGAACATGG 475
QY 553 ATGGGAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGCCGAGCTATA 612
DB 476 ATGGCAAGGAATGTGTAAATGAAACAGGAGGAGTCTCTCCGAGCCACTTCCGCTTACA 535
QY 613 AGCAGATATCTCTCTGTGACACTTACACTACCTCAGAGAGAGAGGAGCCACTTTTATT 672
DB 536 ACACAGAAGTGTGTGTGACCTTACCTTACCTCAGGGAAG-----TTCAGTGTGG 589
QY 673 TTCCAAATACATCAGTTTACCCAGAACGACGCAAGAAATACAAAGGCCAGCATTTAC 732
DB 590 CCTATAACACAACTCCCATACAGACCAAGTCTCTCAGACTGAATGTCAAGGAGACCTT 649
QY 733 CTTATGACAGCAGGAGAGATCAGCGTGAAGAGTCAAGCCATCCCACTCAGTCAAAAG 792
DB 650 CTTATGACTCTGTCTCAGGAGAGGAGCATGGAACAATAATGAACTCTGGCCCTCAACAAA 709
QY 793 CTACCAACCAATCATCTTCA---ACAGTGGCCCAACAGAGACAGCGTCTCAGTTAG 849
DB 710 GTCCCTCTCTTGGAGGATCACAGACCATGTTGGGCAAGAACACTGAGCAGCGGCCAGCCAG 769
QY 850 ATCCCTATCAGTTCTTGGACGACAGCAGCGCTTGTGAAATCCAGGAGTGGGAGA 909
DB 770 ATCCCTATCAGATCTCTGGGCGCAACAGAGCGGCTTAGCAACCTGGGAGTGGGAGA 829
QY 910 TACAGTATGGCAGTCTCTTCTGAGCTTCTGTGGACAGCTTCCAACTTCCAACTGCAATCA 969
DB 830 TCCAGCTGTGGCAGTTCTCTCTGGAAGTACTGTCTCGACAGCGCCAAACCCAGCTGTATCA 889
QY 970 CTGGGAGGGCAAAATGGGAGTTCAAGATCAGACACCTGATGAAGTCTCGGCTT 1029
DB 890 CTGGGAGGGGCAACAGCGGAGTTCAAATGACGACCTGATGAGTGGCCAGGCGCT 949
QY 1030 GGGGAGAGGAGAAAGCAAACTAATCACTATGACAACTCAGCGCTGCACTTCGCT 1089
```

```
950 GGGGAGAGCGGAAGAGCAAGCCCAACATGAATTATGACAAGCTGAGCGGGCCCTCCGAT 1009
QY 1090 ACTACTATGACAAATATATTGACTTAAGTTTCATGGTAAAGCTATGCTTCAACAATTG 1149
DB 1010 ACTACTATGACAAATATATTGACTTAAGTTTCATGGTAAAGCTATGCTTCAACAATTG 1069
QY 1150 ATTTCCAGGAATCGCTCAGGCCCTCCAGCCTTCCAGCTTCCAGAAATCATCCATGTACAAT 1209
DB 1070 ACTTCCATGGCATTGCCAGGCCCTTGACGCCACATCCAAACAGAGACATCCATGTACAAGT 1129
QY 1210 ACCATCAGACCTTCCCTTACATGAGTTCTTCTACCATGCGACACCCCGAGAGATGAACTTTG 1269
DB 1130 ATCCCTCTGATATCTCTTACATGCTTCTTCTACATGCTTCTTCTTCTTCTTCTTCTTCTTCT 1189
QY 1270 TAGCTTCCCATCCCTCTGCTTTGCCCGTAACTTCTTCCAGCTTTTCTGTCGCCCTTAATC 1329
DB 1190 TCCCGTCTTCAACCATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1249
QY 1330 CATCTGGAATTCACCAACTGAGGAGCATCTTCCCAATACCAAGCTGCCAGCTGCTCAT 1389
DB 1250 AATCTGGAACCTTCCCGGACTGCTGGGATCTATCCAAACCCAGTGTCCCGCCCATCTTA 1309
QY 1390 -----TGCTTCCCATCTTGGCACCCTACTACTA 1417
DB 1310 ACACCCAGCTGCTTCCATCTAGGAGCTACTACTA 1346

RESULT 9
AK036655 3118 bp mRNA linear HTC 19-SP-2003
LOCUS Mus musculus adult male bone cDNA, RIKEN full-length enriched
DEFINITION library, clone:9830146N17 product:Friend leukemia integration 1,
full insert sequence.
ACCESSION AK036655.1 GI:26331575
VERSION AK036655
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
```

AUTHORS	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team	
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
JOURNAL REFERENCE	Nature 420, 563-573 (2002)	
AUTHORS	6 (bases 1 to 3118)	
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takanashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	
TITLE	Direct Submission	
	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	
JOURNAL	cdna library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.	
	Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
COMMENT	Please visit our web site for further details.	
	URL: http://genome.gsc.riken.go.jp/ URL: http://phantom.gsc.riken.go.jp/ Location/Qualifiers	
FEATURES	1. .3118	
	/organism="Mus musculus"	
source	/mol_type="mRNA"	
	/strain="C57BL/6J"	
CDS	/db_xref="PANTOM_DB:9830146N17"	
	/db_xref="MGI:2401092"	
	/db_xref="taxon:10090"	
	/clone="9830146N17"	
	/sex="male"	
	/tissue_type="bone"	
	/clone_lib="RIKEN full-length enriched mouse cDNA library"	
	/dev_stage="adult"	
	132..1547	
	/note="unnamed protein product; Friend leukemia integration 1 (MGD MGI:95554, GB NM_008026, evidence: BLASTN, 99%, match=1671) putative"	
	/codon_start=1	
	/protein_id="BAC29518.1"	
	/db_xref="GI:26331576"	
	/translation="MDGTEKEALSIVSDQSLFDSAYGAAALPKADMTASGSPDYQG PHKINLPQQQWVNPVVRVDMGNSRSPVDCVCKNKLVGGEIDTPNFON SYMDKNGPPNTERVTPADPTLWTOEHVROWLEWAIKEYGLMEIDTFPFON YRGKELCKNKKEDFLRATSAVNTVELLSLYRESSLLAYNTSHTDQSSRLANPK PSYSVRRGAWNNMNSGLNKSPLLGSGQTGMKZTEQRPQDPVQILQPTSRLANPK SQCIQLWQFLSDSNANSCITWEGTNGEFTMDPEVARWRGERSKKNMNDKL SVALRYDYDNTIMTVHGRYAYKFDHFHIAQLQPHPTETSMYKPSDIDSYMPSYHA HQQKNVFPVSHPSMPVTSSSPFGAASQWTSFTAGIYFNPSPVPRHNPVSHLGSY Y"	
ORIGIN	38.8%; Score 559; DB 11; Length 3118;	
	Query Match Best Local Similarity 66.8%; Pred. No. 1.5e-155; Matches 917; Conservative 0; Mismatches 420; Indels 36; Gaps 7;	
QY	63 ATGCGACGACTATTAAAGAGCATTATCAGTGGTCAAGACACAGTCCTTGTGTGAG 122	
	192 ATGGACGGGACTATTAAAGAGGCTCTGTCTGTGGTGAAGTGAACATGATGATGATG 251	
Db		
QY	123 TGTGGCTTACGATCG---CCGCACCTTTGCAAAAGACAGAAATGACAGCTCTCTCCAGT 179	
	252 TCAGCATACGGACGCGAGCCCATCTCCCAAGGAGATATGACTGCTTGGGGAGTCTCT 311	
Db		
QY	180 GAATATATGGCAAAACATCAAAGATGAGCCCGCGTTCCTCCAGCAGACGACTGGTTAT 239	
	312 GACTACGGGCGAGCCCAAAAATCAACCCCTCCCAACCGCAGCAGGAGTGGATCAACCC 371	
Db		
QY	240 CCCCCGCGCAGAGTTTACCATTAGATGGAGTGAACCCAAACCCAGGTTAATGGTCAAG 299	
	372 CC-----AGTGAGAGTCAATGTCAAGCGGGAGTATGACCAATGATGATGATGATG 422	
Db		
QY	300 AATTCACTGATGACTGCGAGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTAGACA 359	
	423 GAGTCTCGGTGACTGCACTGT---CAATGTAAACAAGCTGTGTGGCGGAGCGCAAGCC 479	
Db		
QY	360 GTTGGGATGACATATGGAAGCTACATGAGAGAGCA---TATTCCGCTCCCAATATG 416	
	480 AACCCATGAACTATATAGCTAGTGGATGAGAGAGAGCGCCCTCTCTCCACATG 539	
Db		
QY	417 ACAACCAATGAACGAAGAGTTATTGTGCCAGCAGATCTCTACGTTATGGACACAGACC 476	
	540 ACCACCAACGAACGGAGAGTCAATTGTGCTGCGAGACCCCACTGTGGACACAGGAGC 599	
Db		
QY	477 GTAGGCGAGTGGCTGAGTGGCGAGTGAAGAGTATGGTCTCCAGACGTCGACATCTTG 536	
	600 GTTCGACAGTGGCTGAGTGGGCTATAAAGAAATACGGATGTATGGAGATGACACTTC 659	
Db		
QY	537 TTGTTCAGAAACATTGATGGGAAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGA 596	
	660 TTCTTCAGAACATGGATGGCAGAGGATTTGTATAATGAACAAGGAGGACTTCTCCGA 719	
Db		
QY	597 CTCAGCCGAGCTATTAACGAGATATCTCTCTGTCACTTACCTACTACTCTCAGAGAGA 656	
	720 GCCACCTCCGCTTACAAACAGAGAGTCTTGTGCGACCTCAGTTACCTCCTCGGAAAG 778	
Db		
QY	657 GGAGCCACTTTTATTTTCCAAATATCATCATGTTTACCAGAAGCAACGCAAAAGATAACA 716	
	779 -----TTCACTGTGGCTTATAACAACCTCCATACAGACCAGTCTCACGACTGAAT 833	
Db		
QY	717 ACAAGCCAGATTATACCTTATGAGCAAGCGAGAGATACAGCTGGAGAGTCAAGCCAT 776	
	834 GTCAAGGAAGACCTCTTATGACTCTGTCTCAGGAGAGGAGCATGGAACAATAATATGAAC 893	
Db		
QY	777 CCNACTCAGTCAAAGCTACCAACCATCATCTTCA---ACAGTGGCCCAAAACAGAGAC 833	
	894 TCTGGCTCAACAAAAGTCTCTCTCTGGAGGATCACAGACCATGGCAAGACACTGAG 953	
Db		
QY	834 CAGCGTCTCTCAGTTAGATCTTATCAGATTCTTTGGACCGACCAAGCAGCCCTTTGAAAT 893	
	954 CAGCGGCCCCAGCCAGATCTTATCAGATCTCTGGGCGCAACCAAGCAGCCCTTAGCAAA 1013	
Db		
QY	894 CCAGGAGTGGCGAGATACAGCTATGCGAGTCTTCTGAGGATCTGTGCGACAGCTCC 953	
	1014 CTTGGAGTGGCGAGATCCAGCTGTGCGAGTCTTCTCTGGAATCTCTGTCGACAGCGCC 1073	
Db		
QY	954 AACTTCCAACATGCATCACTGGGAGGGCACAAATGGGGAGTTCAGATGACAGACCTGAT 1013	
	1074 AACCCAGCTGTATCACCTGGAGGGGACCAACGGGGAGTTCAAATGACGGACCTGAT 1133	
Db		
QY	1014 GAATGGCTGGCGTTGGGAGAGAGAAAGCAACCTAACATGACTATGACAACTC 1073	
	1134 GAGTGGCCAGGGCGCTGGGAGAGCGGAAAGCAAGCCCAACATGAATTTATGACAGCTG 1193	
Db		
QY	1074 AGCCGTGCATTCGCTACTACTATGCAAAAAATATATGATTAAGTTTCATGTTAAACGC 1133	
	1194 AGCCGGGCCCTCCGATCTACTATGACAAAAACATATATGACCAAGTGCATGCGCAAGG 1253	
Db		
QY	1134 TATGCCCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCTCACCCTCCAGAA 1193	
	1254 TATGCTACAGTTTGACTTCCATGGCATTTGCCAGGCCCTTGCAGGCCACATCCAAACAG 1313	
Db		

```
QY 1194 TCATCAATGTAACAAATACCCATCAGACCTCCCTACATAGTTCCTACCAATGCACACCCC 1253
Db 1314 ACATCCATGTACAGTATCCCTCTGATATCTCTACATGCTTCCTACCAATGCATCAA 1373
QY 1254 CAGAAGATGAATTTGTAGTCCCATCCCTCCCTGTTGCGGTAACCTCATCCAGCTTT 1313
Db 1374 CAGAAGGTGAATTTGTCCCGTCTCACCCATCTCCCATGCTGTCACCTCTCCAGCTTC 1433
QY 1314 TTGTGTCGCCCTAATCCATATCGGAATTCACCAACTGGAGGCATCTACCCCAATACCAGG 1373
Db 1434 TTGGAGCAGATCAATATATGACCTCCCTCCCTGCTGGATCTATCCAAACCCCACT 1493
QY 1374 CTGCGAGTGTCTATA-----TGCTTCCCATCTTGGACCTACTACTA 1417
Db 1494 GTCCCCCGCCATCTTAACACCCACGTGCTTTCACACTTAGGCAGTACTACTA 1546

RESULT 10
AY418893 1347 bp DNA linear GSS 17-DEC-2003
LOCUS Homo sapiens FLI1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY418893
VERSION AY418893.1 GI:39774853
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1347)
AUTHORS Clark,A.G., Ghanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1347)
AUTHORS Clark,A.G., Ghanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1 .1347
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1347
/gene="FLI1"
/locus_tag="HCM6709"

gene

ORIGIN
Query Match 38.2%; Score 552.6; DB 29; Length 1347;
Best Local Similarity 66.3%; Pred. No. 8.5e-154;
Matches 900; Conservative 0; Mismatches 424; Indels 33; Gaps 6;

QY 79 AGGAAGCATTTACAGTGGTGAAGACAGTCCCTCTTCCAGTGAATATGGGCAACAT 195
Db 5 AGGAGGCTCTGTGGTGGTGAAGACAGTCCCTCTTCCAGTGAATATGGGCAACAT 138
QY 139 ---CCACCTTGAAGACAGTGAAGACAGTCCCTCTTCCAGTGAATATGGGCAACAT 195
Db 65 CAGCCCATCTCCCAAGGCGGACATGACTGCTTCGGGGAGTCTGACTAGGCGGCGCC 124
QY 196 CAAAGATGAGCCCGCGTTTCCCGCAGCAGGACTGGTTATCAGAGCCCGGCGCAGTTA 255
Db 125 ACAGATCAACCCCTCCACACACAGCAGGAGTGGATCAATCAGCC-----AGTGA 175
```

```
QY 256 CCATTAAGATGGAGTGTAAACCCAAACCCAGGTTAATGGGTCAAGGAATTCACCTGATGACT 315
Db 176 GGGTCAACGCTCAAGCGGAGTATGACCACATGGAATGGATCCAGGGAGTCTCCGGTGGACT 235
QY 316 GCAGCGTGCAAAAGGAGGGAATGGTTAGCAGTTTCAGACAATGTTGGATGAACATATG 375
Db 236 GCAGCGTTAGCAAAATGCGCAAGCTGGTGGGCGGAGGCGAGTCCAAACCCATGAATACA 295
QY 376 GAAGCTTACATGAAGAGAGAGCAT---ATTCCGCCTCCAAATATGACAACCAATGAACGAA 432
Db 296 ACAGCTATATGACAGAGAGAGTGGCCCCCTCTCCCAACATGACCAACCAACGAGAGGA 355
QY 433 GAGTTATTGTGCGCAGCAGATCTCTAGCTTATGAGCAGACAGCACCATGTAGCGGAGTGGCTGG 492
Db 356 GAGTCATCTCTCCCGCAGACCCCACTGTGGACACAGGAGCATGTGAGGCAATGGCTGG 415
QY 493 AGTGGGCACTGAAGAGTATGGTCTTCCAGACGTGGACATCTTGTGTTTCCAGAACATTTG 552
Db 416 AGTGGGCCATTAAGAGGTACAGCTTGATGGAGATCGACATCTCTTTTCCAGAACATGG 475
QY 553 ATGGGAAAGAGTTGTGTAATAATGACCAAGATGACTTCCAGAGACTCAGCCGAGGTATA 612
Db 476 ATGGCAAGGAACTGTGTAATAATGAACAGGAGGACTTCTCCGCGCACCAACCTCTACA 535
QY 613 ACGCAGATATCTCTCTGTCTACACCTACACTACTCTAGAGAGAGAGGAGCCACTTTTATT 672
Db 536 ACACGGAAGTGTGTTGTCTACACCTCAGTTACTCAGGGAAG-----TTCACGTCTGG 589
QY 673 TTCAAATATCATGAGTTTATCCAGAAAGCAACGCAAGAATAACAAAGCCAGATTTAC 732
Db 590 CCTATAATACAACTCCCAACCGCAATCTCTCAGATTGAGTGTCAAGAAGACCCCTT 649
QY 733 CTTATGACGACGCGAGGATCAGCGTGCAGAGTCACAGCCATCCCACTCAGTCAAAA- 791
Db 650 CTTATGACTCAGTCACAGAGGAGCTTGGGCAATACATGATTTCTGSCCTCAACAAA 709
QY 792 --GCTACCCCAACCATCATCTTCAACAGTGCCTCCAAAACAGAAAGCAGCGTCTCAGTTAG 849
Db 710 GTCTCTCCCTTGGAGGGGCAACAAACATCAGTAAGAATAACAGAGCAACGGCCCCAGCAG 769
QY 850 ATCTTTATCAGATCTTTCAGCCAGCAGCAGCGCTCTTCAAAATCCAGGAGTGGCGAGA 909
Db 770 ATCCGTATCAGTCTCTGGGCCCCACAGCAGTCCCTTAGCCATCCCAACCTTGGAGCGGCGA 829
QY 910 TACAGCTATGGCAGTTCCTACTGAGCTTCTGTGGAAGCTCCAACTCAACATGATCA 969
Db 830 TCCAGCTGTGGCAATTCCTCTGAGCTGCTCTCCGACAGCGCAACGCCAGCTGTATCA 889
QY 970 CTTGGGAGGCGCAAAATGGGAGTTCAAGATGACAGACCTCTGATGAAGTGGCTCGCGCTT 1029
Db 890 CTTGGGAGGCGCAACCGGGAGTTCAAAATGACGACCCCGATGAGTGGCCAGCGCT 949
QY 1030 GGGGAGAGAGAAAAGCAAACTTAACATGAACCTATGACAACTTCAGCCGCTGCACTTCGCT 1089
Db 950 GGGGAGGCGGAAAAGCAAGCCCAACATGAATTAACGACAGCTGAGCGGGCCCTCGCTT 1009
QY 1090 ACTACTATGACAAAATATTATGACTTAAGTTCAATGGTAAACCTATGCTACAAATTTG 1149
Db 1010 ATTACTATGATAAAAACATTTATGACCAAAAGTGCACGCAAAAAGATATGCTTTACAAATTTG 1069
QY 1150 ATTTCCAGGGAATCGCTCAGGCGCTCCAGCTCCTCACTCCAGAAATCATTCATGTACAAAT 1209
Db 1070 ACTTCCACGCAATGCGCCAGCTCTCGAGCCACATCGACCGAGTCTGCTCATGTAAGT 1129
QY 1210 ACCCATCAGACCTCCCTCATATGAGTTCTTACATGACACCCCGCAGAAATGAACCTTTG 1269
Db 1130 ACCCTTCTGACATCTCTCATACCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1189
QY 1270 TAGCTCCCGCATCCCGCTGCTTTGCGGTAACTCTATCCAGCTTTTGTGCTGCTTCTTCTTCTTCT 1329
Db 1190 TCCCTCCCATCCATCTCTCCATGCTGCTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1249
```


QY 1330 CATACTGGAATTCACCAACTGGAGGCAATCTACCCCAATACA-----GGCTGCCAG 1380
 |||||
 Db 1250 AATACTGGACCTCCGCCACCGGGGGAATCTACCCCAACCCCAAGCTGCCCGGCATCCTA 1309
 |||||
 QY 1381 CTGCTCATATGCTTCCCATCTTGGCACCTACTACTA 1417
 |||||
 Db 1310 ACACCCAGTGCTTCACACTTAGGCGAGCTACTACTA 1346
 |||||

RESULT 11
 BU478556/c
 LOCUS 606 bp mRNA linear EST 30-NOV-2002
 DEFINITION 603471895F1 CSEQBN22 Gallus gallus CDNA clone CHEST349m24 5', mRNA
 sequence.
 ACCESSION BU478556
 VERSION BU478556.1 GI:25972133
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 606)
 Fong, W.T., Tickle, C., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Boardman, P.E., Sanz, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 MEDLINE
 PUBMED
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..606
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Layer and broiler"
 /db_xref="taxon:9031"
 /clones="CHEST349m24"
 /sex="Male and female"
 /tissue_type="Chondrocytes isolated from growth plate
 cartilage"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQBN22"
 /note="Vector: pBluescript II KS(+); Site 1: EcoRI;
 Site 2: NotI; This normalized library was constructed from
 1 million independent clones. cDNA synthesis was initiated
 using an oligo(dT) primer, using methylated C in the first
 strand synthesis reaction. Following this first strand
 reaction, double-stranded cDNA was blunted, ligated to
 NotI adapters, digested with EcoRI, size-selected, and
 cloned into the NotI and EcoRI compatible sites of a
 custom modified MCS of the pBluescript (KS+) vector. The
 library was normalized in 2 rounds using conditions
 adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
 Ronaldo et al., Genome Research 6 (1996): 791, except that
 a significantly longer reannealing hybridization was
 used."

ORIGIN
 Query Match 36.4%; Score 526.2; DB 13; Length 606;
 Best Local Similarity 97.5%; Pred. No. 4.8e-146;
 Matches 545; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
 707 AAGATTAACAACAGGGCCAGATTTCCTATGACCAAGCGAGGAGATCAACGCTGGACGAG 766
 |||||
 Db 590 AAGGATTACCAAGGCCAGATTTCCTATGACCAAGCGAGGAGATCAACGCTGGACGAG 531
 |||||

QY 767 TCACAGCCATCCCACTCACTCAGTCAAA-AGCTACCCAAACCATCATCTTTCAACAGTGCACAAA 825
 |||||
 Db 530 TCACAGCCATCCCACTCACTCAGTCAAAAGAGTACCCCAACCATCATCTTTCAACAGTGCACAAA 471
 |||||
 QY 826 CAGAGAGCAGCGTCTCAGTCTAGATCCTTATCAGATTCTTGGACCGACACGACCGTC 885
 |||||
 Db 470 CAGAAGACCAAGCGTCTCAGTCTAGATCCTTATCAGATTCTTGGACCGACACGACCGTC 411
 |||||
 QY 886 TTGCAAAATCCAGGAGTGGGAGATACAGTATGCGAGTTCCTACTGGAGCTTTCTGTGG 945
 |||||
 Db 410 TTGCAAAATCCAGGAGTGGGAGATACAGTATGCGAGTTCCTACTGGAGCTTTCTGTGG 351
 |||||
 QY 946 ACAGTCTCAACTCCAACTGCATCACTGGGAGGACAAATGGGAGTTCAGATGACAG 1005
 |||||
 Db 350 ACAGTCTCAACTCCAACTGCATCACTGGGAGGACAAATGGGAGTTCAGATGACAG 291
 |||||
 QY 1006 ACCCTGATGAAGTGGCTGGCGTTGGGAGAGAGAGAAAAGCAAACTAACATGAACATG 1065
 |||||
 Db 290 ACCCTGATGAAGTGGCTGGCGTTGGGAGAGAGAGAAAAGCAAACTAACATGAACATG 231
 |||||
 QY 1066 ACAACTCAGCGCTGCATCTTCTACTACTATGACAAAAATATTATGATAAAGTTCTATG 1125
 |||||
 Db 230 ACAAACTCAGCGCTGCATCTTCTACTACTATGACAAAAATATTATGATAAAGTTCTATG 171
 |||||
 QY 1126 GTAAAGGCTATGCTTACAAATTTGATTTCCACGGGAATCGCTCAGGCCCTCCAGCTCACC 1185
 |||||
 Db 170 GTAAAGGCTATGCTTACAAATTTGATTTCCACGGGAATCGCTCAGGCCCTCCAGCTCACC 111
 |||||
 QY 1186 CTCGAGAATCATCATGTACAAATACCATCAGAGCTCCCTTACATGATGTTCTTACCATG 1245
 |||||
 Db 110 CTCGAGAATCATCATGTACAAATACCATCAGAGCTCCCTTACATGATGTTCTTACCATG 51
 |||||
 QY 1246 CACACCCCCAGAGATGAA 1264
 |||||
 Db 50 CACACCCCCCAAAAAA 32
 |||||

AY418894 1256 bp DNA linear GSS 17-DEC-2003
 Pan troglodytes FLII gene, VIRUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY418894
 VERSION AY418894.1 GI:39774854
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 1 (bases 1 to 1256)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)
 14671302
 2 (bases 1 to 1256)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 Direct submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 Location/Qualifiers
 1..1256
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"

gene	ORIGIN	Query Match	Best Local Similarity	Mismatches	Conservative	Score	DB 29;	Length	DB 29;	Indels	DB 29;	Gaps	DB 29;
		/db_xref="taxon:9598"	34.6%;	Score 501.2;	DB 29;	Length 1256;							
		<1..>1256	66.2%;	Pred. No. 2.2e-138;									
		/gene="FLI1"	0;	Mismatches 376;	Indels	21;	Gaps	4;					
		/locus_tag="HCM6709"											
		ORIGIN											
		Query Match											
		Best Local Similarity											
		Mismatches 777;											
		Conservative											
		Score											
		DB 29;											
		Length											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											
		Gaps											
		DB 29;											
		Indels											

Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHAM10420 row: f column: 18
 High quality sequence stop: 593.
 Location/Qualifiers
 1. .715
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4521233"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_hosts="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_92"
 /note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

FEATURES

source

ORIGIN

Query Match	33.3%	Score 482.2;	DB 12;	Length 715;
Best Local Similarity	84.1%	Pred. No. 8.2e-133;		
Matches	544;	Conservative	0;	Mismatches 103; Indels 0; Gaps 0;

Qy	8	CGCAACGATTAATTTATTAGCAATTTATTAGCATCAATAATCTTGATCACATTATGGC	67
Db	39	GAGAAGCAATAATTTATTAAACATTTATTAAACATCAATAAATGATTGCATTATGGC	98
Qy	68	AAGCACTATTAAAGGAAGCATTTATCAGTGGTGAAGACCAGTCCTTGTGAGTGTGC	127
Db	99	CAGCACTATTAAAGGAGCCATTATCAGTTGTGAGTGAGGACCAGTCGTTGTTGAGTGTGC	158
Qy	128	CTACGGATCGCCCAACCTTGCAAGACAGAAATGACAGCCCTCCTCCAGTGAATATGG	187
Db	159	CTACGGAAACGCCACACACCTGGTAAACAGAGATGACCGCGTCCTCCTCCAGCGACTATGG	218
Qy	188	GCMAACATCAAGATGAGCCCGCGTTCCGCCAGCAGACTGTTATCACAGCCCCCGGC	247
Db	219	ACAGACTTCCAGATGAGCCACCGCTCCCTCAGCAGGATGGTGTCTACACCCCGAGC	278
Qy	248	CAGAGTTACCATTAAGATGAGTGTAAACCCAAACAGGTTAATGGGTCAAGGAATTCACC	307
Db	279	CAGGTCACCATCAAAATGGAATGTAACTTAGCCAGGTGAATGGCTCAGGAATCTCC	338
Qy	308	TGATGACTGCAGCGTGGCAAGAGGAGGAAATGTTAGCAGTTCAGACATGTTGGGAT	367
Db	339	TGATGAATGCAGTGTGGCCAAAGGCGGGAAGATGGTGGCAGCCCGAGACACCGTGGGAT	398
Qy	368	GAACTATGGAAGCTACATGGAAGAGAGACATATTCGGCTCCAAATATGACAAACCAATGA	427
Db	399	GAACTACGGCAGCTACATGGAGGAGAGACATGCCACCCCAACATGACCAGGAACGA	458
Qy	428	ACGAAGAGTTATTGTGCCAGCAGATCCTACGTTATGGAGCAGACAGCCATGTACGGCAGTG	487
Db	459	GCCGAGAGTTATCGTGCCAGCAGATCCTACGCTATGGAGTACAGACCATGTGCGGCGAGTG	518
Qy	488	GCTGAGTGGGCGAGTGAGAGGATATGGTCTTCCAGACGTGGACATCTTGTGTTCCAGAA	547
Db	519	GCTGAGTGGGCGGTGAAGATATGGCTTCCAGCGTCAACATCTTGTGTTATCCAGAA	578
Qy	548	CATTGATGGGAAGAGTTGTGTAAATGACCAAAGATGACTTCCAGACTCAGCCGAG	607
Db	579	CATCGATGGGAAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCAACCCGAG	638
Qy	608	CTATAACGAGATATCCTCCTGTCCACACTACCTACCTCAGAGAGA	654
Db	639	CTACACGCCGACATCCTTCTCTCATCTCTCAATCTCCTCAGAGAGA	685

Search completed: May 27, 2004, 17:47:59
 Job time : 4104 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 27, 2004, 19:27:37 ; Search time 5617 Seconds
(without alignments)

3480.098 Million cell updates/sec

US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTIKALSVDQSLFE.....IYNTRLPAAMPSHLGTYV 451

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n_model -DEV=xlh
-Q=/cgn2.1/USPRO.spool/US0902772/runat_27052004.162841.2144/app_query.fasta_1.647
-DB=GenEmbl -QFWT=rascap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0902772@cgn 1.1 3731 @runat_27052004.162841.2144 -NCPV=6 -ICPV=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.scs.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rtd.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	2444	100.0	1413	5	AY065661 Gallus ga
2	2444	100.0	1447	6	E31253 Protein hav
3	2414.5	98.8	1516	5	X77159 G.Gallus ER
4	2414.5	98.8	1528	6	E31254 Protein hav
5	2327.5	95.2	1509	6	AX657228 Sequence
6	2313	94.6	2016	9	BC040168 Homo sapi
7	2287	92.8	2209	10	AB073080 Mus muscu
8	2192	89.7	2012	5	AJ224126 Xenopus l
9	2182	89.3	3166	9	M17254 Human erg2
10	2179	89.2	1808	10	AB031088 Rattus no
11	2159.5	88.4	2133	10	AB073079 Mus muscu
12	2153.5	88.1	1798	5	AJ224125 Xenopus l
13	2151	88.0	2131	10	AB073078 Mus muscu
14	1955.5	80.0	1391	6	AX657230 Sequence
15	1946	78.8	1372	6	AX657234 Sequence
16	1840	74.1	3126	9	HUMERG11
17	1807.5	74.0	1219	6	AX657232 Sequence
18	1637	67.0	1436	5	DRE249590 Danio rer
19	1624.5	66.5	1359	9	AX695583 Sequence
20	1624.5	66.5	1359	9	AY029368 Homo sapi
21	1624.5	66.5	2394	9	BC010115 Homo sapi
22	1624.5	66.5	2316	6	A36461 Sequence 2
23	1624.5	66.5	2398	6	AR080101 Sequence
24	1624.5	66.5	2398	9	X67001 H.sapiens H
25	1624.5	66.5	2397	6	AX587825 Sequence
26	1624.5	66.5	2397	6	AX695582 Sequence
27	1624.5	66.5	2397	9	AX695582 Sequence
28	1624.5	66.5	2359	9	HUMERGBFLI
29	1624.5	66.5	6849	12	BC001670 Homo sapi
30	1621.5	66.3	1673	9	AY029367 Synthetic
31	1613.5	66.0	1359	6	S45205 Fli-1-frien
32	1613.5	66.0	1729	6	AX695580 Sequence
33	1613.5	66.0	1729	6	AX695579 Sequence
34	1607	65.8	2919	5	X59421 Mouse Fli-1
35	1599	65.4	2854	6	BC056588 Danio rer
36	1579.5	64.6	3450	5	BD094081 Shear str
37	1576	64.1	4403	5	Y14773 Coturnix co
38	1567.5	64.1	3545	5	X66579 X.laavis mr
39	1536.5	62.9	1932	9	Y14774 Coturnix co
40	1486.5	60.8	1245	5	M93255 Human Fli-1
41	1245	51.0	2722	5	AF177538 Danio rer
42	1227	50.2	816	10	BC055627 Danio rer
43	1183.5	48.4	762	9	S66169 Erg-3-immun
44	965	39.5	1431	9	S72621 EMS - erg
45	956.5	39.1	14518	9	AF327066 Homo sapi
C					AP001423 Homo sapi

ALIGNMENTS

RESULT 1

AY065661 1413 bp mRNA linear VRT 17-DEC-2001
 Locus Gallus gallus erg isoform C-1-1 mRNA, complete cds; alternatively spliced.
 DEFINITION
 ACCESSION AY065661
 VERSION AY065661.1 GI:17887440
 KEYWORDS
 SOURCE Gallus gallus (chicken)
 ORGANISM
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 1413)
 AUTHORS Iwamoto, M., Higuchi, Y., Koyama, E., Enomoto-Iwamoto, M., Kurisu, K., Yeh, H., Abrams, W.R., Rosenbloom, J. and Pacifici, M.
 TITLE Transcription factor ERG variants and functional diversification of chondrocytes during limb long bone development
 JOURNAL J. Cell Biol. 150 (1), 27-40 (2000)
 MEDLINE 20351415
 PUBMED 10893254
 REFERENCE 2 (bases 1 to 1413)
 AUTHORS Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K., Koyama, E., Yeh, H., Rosenbloom, J. and Pacifici, M.
 TITLE The role of ERG (ets related gene) in cartilage development
 JOURNAL Osteoarthritis. Cartil. 9 Suppl A, S41-S47 (2001)
 MEDLINE 21535378
 PUBMED 11680687
 REFERENCE 3 (bases 1 to 1413)
 AUTHORS Iwamoto, M., Higuchi, Y., Yeh, H. and Pacifici, M.
 TITLE Direct Submission
 JOURNAL Submitted (05-DEC-2001) Oral Anatomy and Developmental Biology, Osaka University Faculty of Dentistry, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan
 FEATURES
 source Location/Qualifiers
 1..1413
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /db_xref="taxon:9031"
 37..1132
 /note="alternatively spliced; transcription factor"
 /codon_start=1
 /product="erg isoform C-1-1"
 /protein_id="AA140889.1"
 /db_xref="GI:17887441"
 /translation="VASTIKELSVSEDSQIFRCAYGSPHLAKTEMTASSSEYGGT SKSPRPVQQWLSQPPARVTIKMECPNQNGSRNSDDCVAGKRMVSSSDNVGM NYGSMEEKHIPPNNMTNRRVIVPADPTLWSTDRVQWLEWAKVYGLPDVDILF QNIDGKELCKTKDFQLRTSPYNADILLSHLYLRERGAFTI FPNSTVYPEATQRTI TRPDLPYQARRSAMTSHTSQSKATQSSSTVPKTEDQRPQDLPYQILPQTSRLA NPGSQIOLWQFLLELSDSNSCITWEGTNGRFPKTPDDEVARRWGRKSKPNMY DKLSALRYVDKNTMTVKHGRVAYKEDFHGIAQLQPHPPSSMYKPSDLPYMS YAHAPQKNFVAPHPALPVPVTSSEFPAPNPYNSPTGGIYPNTELPAAHMPSHLGY Y"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,02e-162 Length: 1413
 Score: 2444.00 Matches: 451
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 5 Gaps: 0
 US-09-902-772-2 (1-451) x AY065661 (1-1413)
 QY 1 MetAlaSerThrIleIysGluAlaLeuSerValSerGluAaspGlnSerLeuPheGlu 20
 DB 37 ATGGCAAGACATTATGAAGACATTATCAGTGGTGAAGACCATCTCTGTTGAG 96
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
 DB 97 TGTGCTACGGATCGCCCACTTCGACAGACAGAAATGACAGCCTCTCTCCAGTGA 156

QY 41 TyrGlyGlnThrSerIysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
 DB 157 TATGGCAAAATCAAGATGAGCCGCGCTTCCACGACGACTGGTTATCAGACCC 216
 QY 61 ProAlaArgValThrIleIysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 DB 217 CCGGCCAGAGTTTACCATTAAGATGGAGTGAACCCAAACAGGTTAATGGGTCAAGGAAT 276
 QY 81 SerProAspAspCysSerValAlaLysGlyIleValMetValSerSerSerAspAsnVal 100
 DB 277 TCACTGATGATCTGACGCTGCAAAAGAGGAAATAGTTAGCAGTTACACATGTT 336
 QY 101 GlyMetAsnTyrGlySerTyrMetGluIleIysHisIleProProProAsnMetThrThr 120
 DB 337 GGGATCAACTATGGAAGCTACATGGAAGAGAGACATATTCGCCTCCCAATATGCAACC 396
 QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
 DB 397 AATGAACGAGAGTATTATGTGCAGCAGATCTAGTTAAGGACACAGACCATGTACGG 456
 QY 141 GlnTrpLeuGluTrpAlaValIysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
 DB 457 CAGTGGCTGGAGTGGGACGTGAAGGAGTATGTCTTCCAGACGTGGACATCTTGTGTTTC 516
 QY 161 GlnAsnIleAspGlyIysGluLeuCysIysMetThrLysAspAspPheGlnArgLeuThr 180
 DB 517 CAGAACATTGATGGGAAAGAGTTGTGTAATAATGACCAAGATGACTTCCAGAGACTCAG 576
 QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAla 200
 DB 577 CCGAGCTATACCGACATATCTCTGTCCACACTACACTACTCTCAGAGAGAGAGAGCC 636
 QY 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArg 220
 DB 637 ACTTTTATTTTCCAAATACATCAGTTTACCAGAGCAACCAAGATAACAACAAGG 696
 QY 221 ProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHisProThr 240
 DB 697 CCGATTATACCTTATGAGCAGGAGGAGATCAGCGTGGACGAGTCCACAGCCATCCACT 756
 QY 241 GlnSerIleAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260
 DB 757 CAGTCAAAAGCTACCCCAACCATCATCTTCAACAGATGCCCCAAACAGAAACCCAGCGCTCT 816
 QY 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
 DB 817 CAGTTAGATCCCTTATCAGATTCTTGGACCGACGAGCGCGTCTTCCAAATCCAGGAGT 876
 QY 281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsn 300
 DB 877 GGGCAGATACAGCTATGGCAGTCTCTACTGGAGCTTCTGTCGACAGCTCCCACTCAAC 936
 QY 301 CysIleThrTrpGluGlyThrIleGlyGluPheIleValMetThrAspProAspGluValAla 320
 DB 937 TGCATCACCTGGGAGGGGCAAAATGGGAGTTCAAGATCAGACACCTCTGATGAGTGGCT 996
 QY 321 ArgArgTrpGlyGluArgIysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
 DB 997 CGCGCTTGGGAGAGAGAGAAAGCAACCTAATCAATGACTATCACAACCTCAGCCGTGCA 1056
 QY 341 LeuArgTyrTyrThrAspLysAsnIleMetThrLysValHisGlyLysArgGlyAlaTyr 360
 DB 1057 CTTTCGTACTACTATCACAATAATATATGACTAAAGTTTCATGGTAAACGCTATGCTTAC 1116
 QY 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerSerMet 380
 DB 1117 AANTTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCACCTCCAGAAATCATCATG 1176
 QY 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
 DB 1177 TACAAATACCCATCAGACCTCCCTCATGAGTTCTTCCATGACACACCCCGAGAGATG 1236
 QY 401 AsnPheValAlaAlaProHisProProAlaLeuProValThrSerSerSerPheAlaAla 420

```

Db      1237 AACTTGTAGTCGCCATCCCTGCTTTGCGCGTAACCTCATCCAGCTTTTGTGCTGCC 1296
QY      421 ProAsnProTyrTrpAsnSerProThrGlyGlyLeuTyrProAsnThrArgLeuProAla 440
Db      1297 CCTAATCATCTAGTGAATTCACCACTGGAGGCATCTACCCCAATACAGGCTGCCAGCT 1356
QY      441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db      1357 GCTCATATGCTTCCCATCTTGGCACCCTACTACTAC 1389

RESULT 2
E31253
LOCUS   E31253
DEFINITION Protein having cell calcifying inhibitory activity and gene
          encoding the same.
ACCESSION E31253
VERSION   E31253.1 GI:13025685
KEYWORDS JP 199075871-A/1.
SOURCE   unidentified
          unclassified.
REFERENCE 1 (bases 1 to 1447)
AUTHORS  Hiroyasu, I., Yoshinobu, H., Marjio, P., Joel, R. and Helena, E.
TITLE     Protein having cell calcifying inhibitory activity and gene
          encoding the same
JOURNAL   Patent: JP 199075871-A 1 23-MAR-1999;
          CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
          PENNSYLVANIA
COMMENT   OS Unidentified
          PN JP 199075871-A/1
          PD 23-MAR-1999
          PF 29-MAY-1998 JP 1998166076
          FR 18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PI
          HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIUJO PASHIFIKI, PI JOEL
          ROZENBUROOM,
          PI HELENA E
          PC C12N15/09, A61K48/00, C07K14/465, C07K16/18, C12Q1/68 //A61K38/00,
          PC C12P21/02,
          PC C12P21/08, C12N15/00, A61K37/02
          CC Strandedness: Single;
          CC Topology: Linear;
          FH Key Location/Qualifiers
          FT source 1..1447
          FT location/Qualifiers
          /organism='Unidentified'.
          /organism='unidentified'
          /mol_type='genomic DNA'
          /db_xref='taxon:32644'

ORIGIN

Alignment Scores:
Pred. No.: 2,08e-162 Length: 1447
Score: 2444.00 Matches: 451
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-902-772-2 (1-451) x E31253 (1-1447)

QY      1 MetAlaSerThrIleIysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
Db      63 ATGGCAAGCACTATTAAAGGAAGCATTTATCAGTGGTGGAGTGAAGACCACTCTTGTGTGAG 122
QY      21 CysAlaTyrGlySerProHisLeuAlaIysThrGluMetThrAlaSerSerSerGlu 40
Db      123 TGTGCTCATCGGACGCGCCCACTTGGAAAGACAGAAATGACAGCCTCTCTTCCAGTGAA 182
QY      41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db      183 TATGGCAACATCAAGATGAGCGCGCTTCCCGACGAGCACTGGTTATCATCAGGCC 242

```

```

QY      61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db      243 CCGGCCAGAGTTTACCATTAAGATGGAGTGTAAACCAACCCAGGTTAATGGTCAAGGAAT 302
QY      81 SerProAspAspCysSerValAlaIysGlyGlyLysMetValSerSerSerAspAsnVal 100
Db      303 TCACCTGTAGTACCTGCGAGCGTGGCAAAAGAGGAGGAAATGTTAGCAGTTTCAGACAATGTT 362
QY      101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
Db      363 GGGATGAACCTATGGAGAGCTACATGGNAGAGAGCATATTCCGCTCCCAATATGACAACC 422
QY      121 AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
Db      423 AATGAACGAAGAGATTATTGTGCCAGCAGATCCTACGTTTATGGAGCAGACAGACCATGTACCG 482
QY      141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
Db      483 CAGTGGCTGGAGTGGCAGTGAAGAGTATGGTCTTCCAGACGTGGACATCTTGTGTTC 542
QY      161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db      543 CAGAACATTGATGGGAAGAGATTGTCTAAATGACCAAGATGACTTCCAGAGACTCACG 602
QY      181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyVala 200
Db      603 CCGAGCTATAACGAGATATCTCTCTCACCTACACTACTCTCAGAGAGAGAGAGGCC 662
QY      201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArg 220
Db      663 ACTTTTATTTTCCAAATACATCAGTTTACCCAGAGAACCGCAAGAATAACACAAGG 722
QY      221 ProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHisProThr 240
Db      723 CCAGATTACCTTATGAGCAAGGAGAGATCAGCTGGAGTGCAGAGTCCAGCCATCCACT 782
QY      241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260
Db      783 CAGTCAAAAGCTTACCCACCATCATCTTCAACAGTCCCAAAACAGACAGACCGCTCCT 842
QY      261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
Db      843 CAGTTAGATCCTTATCAGATTCTTGGACCGACAGACGCGCTTTCGCAATCCAGGAGT 902
QY      281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuSerSerSerSerSerSerSerAsn 300
Db      903 GGGCAGATACAGCTATGGCAGTTCCTACTGGAGCTTCTCTGGACAGCTCCACTCCAAC 962
QY      301 CysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320
Db      963 TGCATCACCTGGAGGGGCACAAATGGGGAGTTCAAGATGACAGACCGCTGATGAAGTGGCT 1022
QY      321 ArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
Db      1023 CGGCGTTGGGAGAGAGGAAAGCAACCTTAACATGAATATGACAACTCAGACCGGTGCA 1082
QY      341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
Db      1083 CTTTCGCTACTACTATGCAAAAATATATGACTTAAAGTTTCATGTTAAACGCTATGCTAC 1142
QY      361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMet 380
Db      1143 AAATTTGATTTCCACCGAATCGCTCAGCGCCTCCAGCCTCACCTCCAGAAATCATCATG 1202
QY      381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
Db      1203 TACAAATACCCATCAGACCTCCCTCATGAGTTCCTACCATGCACACCCCGAGAAAGATG 1262
QY      401 AsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaA 420
Db      1263 AACTTTGTAGTCCCATCCCTGCTTTGCCCGTAACCTTCATCCAGCTTTTGTGCTGCC 1322

```

ORIGIN					
IWNSTFGGGLIFNIKLPFAAHWFSLHLLII*					
Alignment Scores:					
Pred. No.:	2,57e-160	Length:	1515		
Score:	2414.50	Matches:	450		
Percent Similarity:	94.14%	Conservative:	0		
Best Local Similarity:	94.14%	Mismatches:	1		
Query Match:	98.79%	Indels:	27		
Ds:	5	Gaps:	1		
US-09-902-772-2 (1-451) x GGERG (1-1516)					
QY	274	ArgLeuAlaAsnProGlySerGlyValIleLeuPheMetThrLysMet	CGTCTTCAAATCCAGGGAGTGCGCAGATACACTATGGCAGGTCTCTACTGGAGCTTCG	1022	
DB	963				
QY	294	SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsndlyGluPhelysMet	TGGACACAGCTCCAACCTCCAACCTGCATCACCTGGGGGGGCACAACAAATGGGAGTTTCAAGATG	1082	
DB	1023				
QY	314	ThrAspProAspGluValAlaAlaArgArgTrpGIgLuArgLysSerLysProAsnMetAsn	ACAGACCCTGTGATGAAGTGGCTTCGGCGTTGGGGAGAGAGGAAAAGCAAACCTACATCATGAC	1142	
DB	1083				
QY	334	TyrAspLysLeuSerArgAlaLeuArgTyTYrtyrAspLysAsnIleMetThrLysVal			353

Db 1143 TATGACAAACTCAGCCGTCGCTACTACTATGACAAAAATATATGACTAAGTT 1202

Qy 354 HisGlyLysAspGlyTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373

Db 1203 CATGTAACGCTATGCTCCACAAATTTGATTTCCACGAATCGCTCAGCCCTCAGCCT 1262

Qy 374 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393

Db 1263 CACCCCTCCAGATCATCATGACAAATACCAATCAGACCTCCCTCATAGATTCCTAC 1322

Qy 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProValThr 413

Db 1323 CATGCACACCCCAAGATGAATTTGTAGCTCCCATCCCTGCTTCCCGCTAAC 1382

Qy 414 SerSerSerPhePheAlaProAsnProTyrTyrAsnSerProTyrGlyIleTyr 433

Db 1383 TCATCCAGCTTTTGTGCCCCATCATCTAGCTGAATTCACCAATCGGAGGATCTAC 1442

Qy 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451

Db 1443 CCCATACCGCTCCAGCTGCTCATATGCTTCCCATCTTGGCACCTACTAC 1496

RESULT 4

E31254

LOCUS

DEFINITION

Protein having cell calcifying inhibitory activity and gene encoding the same.

ACCESSION

E31254

VERSION

E31254.1 GI:13025686

KEYWORDS

JP 1999075871-A/2.

SOURCE

unidentified

ORGANISM

unclassified.

REFERENCE

1 (bases 1 to 1528)

AUTHORS

Hiroyasu, I., Yoshinobu, H., Marijo, P., Joel, R., and Helena, E.

TITLE

Protein having cell calcifying inhibitory activity and gene encoding the same

JOURNAL

Patent: JP 1999075871-A 2 23-MAR-1999;

CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA

COMMENT

OS Unidentified

PN JP 1999075871-A/2

PD 23-MAR-1999

PF 23-MAY-1998 JP 1998165076

PR 18-JUN-1997 US 08/978177, 20-JUN-1997 US 60/050297 PI

HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJO PASHIFIKI, PI JOEL ROZENBUROOM,

PI HELENA E

PC C12N15/09, A61K48/00, C07K14/465, C07K16/18, C12Q1/68//A61K38/00,

PC C12P21/02, C12N15/00, A61K37/02

PC C12P21/08, C12N15/00, A61K37/02

CC Strandedness: Single;

CC Topology: Linear;

FH Key

FT Location/Qualifiers

FT source 1..1528

FT /organism='Unidentified'.

FT Location/Qualifiers

FT 1..1528

FT /organism='unidentified'

FT /mol_type='genomic DNA'

FT /db_xref='taxon:32644'

ORIGIN

Alignment Scores:

Pred. No.: 2,6e-160 Length: 1528

Score: 2414.50 Matches: 450

Percent Similarity: 94.14% Conservative: 0

Best Local Similarity: 94.14% Mismatches: 1

Query Match: 98.79% Indels: 27

DB: 6 Gaps: 1

US-09-902-772-2 (1-451) x E31254 (1-1528)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20

Db 63 ATGGCAAGCAGCTATTAAAGAAAGCAATTATCAGTGGTGAGTGAAGACCGAGCTCTTGTTGAG 122

Qy 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40

Db 123 TGTGCTACGGATCGCCACCTTGCAAAGACAGAATGACAGCCTCTCTCCAGTGAA 182

Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60

Db 183 TATGGCAACATCAAGATGAGCCGCGCTTCCCGAGCAGACTGGTTATCACAGCCC 242

Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80

Db 243 CCGGCCAGAGTTACCATTAAGATGAGTGTAAACCAACAGGTTAATGGGTCAAGGAAT 302

Qy 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100

Db 303 TCACCTCATGACTGCGAGCTGGCAAAAGGAGGAAAATGGTTAGCAGTTTACAGCAATGTT 362

Qy 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120

Db 363 GGGATGAACATATGGAAGCTACATGGAGAGAGCATATTCGCGCTCCAAATATGACAACC 422

Qy 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140

Db 423 AATGAACGAAGAGTTATTGTGCCAGCAGATCCTTACGTTTATGGAGCAGACAGACCATGACGG 482

Qy 141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160

Db 483 CAGTGGCTGGAGTGGCAGTGAAGAGTATGGTCTTCCAGAGCTGGACATCTTGTGTTTC 542

Qy 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180

Db 543 CAGAACATTTGATGGGAAAGAGTTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGC 602

Qy 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArg 198

Db 603 CCGAGCTATAACGAGATATCTCTCTGTACACCTACATCTCTCAGAGAGACTCTCTCT 652

Qy 198 ----- 198

Db 663 CCACATTTGACTTCAGATGATGTTGATAAGGCTTACAAAACCTCTCCACGGTTAATGCAT 722

Qy 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213

Db 723 GCTAGAAAACACAGGAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAGCA 782

Qy 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaAlaArgSerAlaTyr 233

Db 783 ACGCAAGATATACAAACAGGCCAGATTTACCTTATGACCAAGCAGGAGATCAGCGTGG 842

Qy 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 253

Db 843 ACGAGTCACAGCCATCCCACTCAGTCAAAAGCTTACCAACCATCATCTTCAACAGTGCCC 902

Qy 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273

Db 903 AAAACAGAGACAGCGCTCTCTAGTTAGATCTTTATCAGATCTTTGACCGCAGCAGCAGC 962

Qy 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeu 293

Db 963 CGTCTTGCAAAATCCAGGAGTGGCAGATACAGCTATGCGAGTTCTCTACTGGAGCTTCTG 1022

Qy 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313

Db 1023 TCGGACAGCTCCCAACTCCCACTGATCAGCTGGAGGGGACAAATGGGGAGTTCAAGATG 1082

Qy 314 ThrAspProAspGluValAlaArgTyrGlyGluArgLysSerLysProAsnMetAsn 333

Db 1083 ACAGACCTGATGAAGTGGCTGGCGTGGGAGAGAGGAAAGCAAACTTAACATGAAC 1142

Qy 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 353

```

Db 1143 TATGACAAACTCAGCGGTGACCTTCGCTACTACTATGACAAAATATATGACATAAGTT 1202
Qy HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
Db 1203 CATGGTAAACGCTATGCTACAAATTTGATTCACCGAATCGCTCAGGCGCTCCAGCCT 1262
Qy HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393
Db 1263 CACCGCTCAGAAATCATCATGATACAAATACCATCAGACCTCCCTCATCATGAGTTCTTAC 1322
Qy HisAlaHisProGlnLysMetAspPheValAlaProHisProAlaLeuProValThr 413
Db 1323 CATGCACACCCCGAGAGATGAACTTTGTAGTCCCATCCCTGCTTTGCCGTAACC 1382
Qy SerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyr 433
Db 1383 TCATCCAGCTTTTGTGCTGCCCTTAATCATCATCTGGAATTCACCACTGGAGGCATCTAC 1442
Qy ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1443 CCCAATACAGGCTGCCAGCTGCCTATGCTTCCCATCTTGGCACCTACTAC 1496

RESULT 5
AX657228
LOCUS AX657228 1509 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 1 from Patent WO03000724.
ACCESSION AX657228
VERSION AX657228.1 GI:29160006
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Luyten, F., de Bari, C. and Dell'Accio, F.
Polynucleotide sequences and vectors useful for the prevention or
treatment of bone- or cartilage-related disorders
Patent: WO 03000724-A 1 03-JAN-2003;
Tigenix N.V. (BE)
FEATURES
Location/Qualifiers
1..1509
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
67..1446
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:29160007"
/db_xref="CAD79711.1"
/translation="MIQTVDPAAHKEALSVMSEDSLSFECAYGTPHLAKTENTASS
SSDYQGTSMGSRPVQDWLSQPPARVTIKMCPNSQVNSRNSPDECSVAKGKRWG
SPDVTGMVNGSVMEKEMPPNMVTTNERRVIVPADPTLWSDHVRQWLEWAVEYGLP
DVAIILFONIDGKEICKMKDDDFDLTPSYNADILLSHLVLRERGAFTFPNTSVYP
EATQRIITRPDLPPPEPRRSATWGHTPOSKAQPSPTVPTKEDQRLDPQYLL
GPTSRLANPQSGQLWQFLLELLSDSSNSCITWEGTNGEFKMTDFDVSARWGER
KSKPNMNYDKLSRALRYDYDKNIMTKVHGKRYAFDFPHGAQALQHPPESSLYKIP
SDLPYMGSHYAHPOKQNFVAPHPALPVTSSFFFAAPNPYNSPTGGIYPNTRLPYTH
NPSHLGLTYV"
ORIGIN
Alignment Scores:
Pred. No.: 3,248-154 Length: 1509
Score: 2327.50 Matches: 427
Percent Similarity: 96.90% Conservative: 10
Best Local Similarity: 94.68% Mismatches: 13
Query Match: 95.23% Indels: 1
DB: 6 Gaps: 1

US-09-902-772-2 (1-451) x AX657228 (1-1509)
Qy 2 AlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerIeuPheGluCys 21

```

```

Db 91 GCAGCTCATATCAAGGAAGCCCTTATCAGTTGTGAGTGAGGACCACTCGTTGTTGAGTGT 150
Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41
Db 151 GCCTACGGAACGCCACACCTGGCTAAGACAGAGATGACCGGCTCTCTCTCCAGCGCATAT 210
Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61
Db 211 GGACAGACTTCCAAGATGAGCCCGCTCTCTCAGCAGGATTTGGTCTCTCAACCCCA 270
Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db 271 GCCAGGGTCACCATCAAAATGGAATGTAACTCCAGGTTGAATGGTCAAGGAACTCT 330
Qy 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGly 101
Db 331 CCGTGAATGATGAGTGTGGCCAAAGCGGGAAGATGGTGGGCGAGCCAGACACCGTTGGG 390
Qy 102 MetAsnTyrGlySerTyrMetGluGlyLysIleProProProAsnMetThrThrAsn 121
Db 391 ATGAACACTACGGCAGCTACATGGAGGAGGACACATGCCCCCAACATGACACCGAAC 450
Qy 122 GluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
Db 451 GAGCGCAGAGTTATCTGTCAGCAGATCTACGCTATGGAGTACAGACCATGTGGCGAG 510
Qy 142 TrpLeuGluTyrAlaValLysGlyTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
Db 511 TGGCTGGAGTGGGCGGTGAAGAAATATGGCTTCCAGACGTCACATCTTTGTTATTCAG 570
Qy 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlnArgLeuThrPro 181
Db 571 AACATCGATGGGAAGAACTGTGCAAGATGACCAAGGAGACTTCCAGAGGCTCACCCC 630
Qy 182 SerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
Db 631 AGCTACAAACGCCGACATCTCTCTCACATCTCCACTACTCTCAGAGAGAGGGTGCAGCT 690
Qy 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
Db 691 TTTATTTTCCAAATACTTTCAGTATATCTGAAGTACGCAAGAAATTAACAATCAGGCCA 750
Qy 222 AspLeuProTyrGluGlnAlaArgArgSerAlaTyrThrSerHisSerHisProThr 240
Db 751 GATTACCATATGAGCCCCCAGGAGATCAGCTTGACCGGTCACGGCCACCCACGCCGCC 810
Qy 241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260
Db 811 CAGTCGAAAGCTGCTCAACCATCTCTTCCACAGTGCCTCAAACTGAAGACCGAGCTCT 870
Qy 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerArgLeuAlaAsnProGlySer 280
Db 871 CAGTTAGATCTCTTATCAGATTCTTGACCAACAAGTAGCCGCTTGCAATTCAGCAGT 930
Qy 281 GlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsn 300
Db 931 GGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGTCTCTGCGACAGCTCCAACCTCCAG 990
Qy 301 CysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320
Db 991 TGCATCAGCTGGGAAGGCCCAACCGGGAGTTCAAGATCAGGATCCCCAGCAGGTGGCC 1050
Qy 321 ArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
Db 1051 CGGCGCTGGGAGAGCGGAGAGCAAAACCCACATGACTAGCTAGCTCAGCCGCGCC 1110
Qy 341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
Db 1111 CTCGGTTACTTACTATGACAAGAACATCATGACCAAGTCCATGGAGCGCTACGCTAC 1170
Qy 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMet 380

```

Db 1171 AAGTTCGACTTCACGGGATCGCCAGGCGCTCCAGCCCCACCCCGGAGTCATCTCTG 1230

Qy 381 TyrIysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400

Db 1231 TACAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATCACGCCACCCACAGAGATG 1290

Qy 401 AsnPheValAlaProHisProAlaLeuProValThrSerSerPhePheAlaAla 420

Db 1291 AACTTTGTGGGCCCCACCTCCAGCCCTCCCGGTGACATCTCCAGTTTITGCTGCC 1350

Qy 421 ProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAla 440

Db 1351 CCAAACCCCACTGGAATTCACCAACTGGGGGTATATACCCCAACACTAGGCTCCCCACC 1410

Qy 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451

Db 1411 AGCCATATGCTCTCTCATCTGGGCACCTACTAC 1443

RESULT 6

BC040168

LOCUS

DEFINITION Homo sapiens v-ets erythroblastosis virus E26 oncogene like (avian), transcript variant 1, mRNA (CDNA clone MGC:49020 IMAGE:6052140), complete cds.

ACCESSION BC040168

VERSION

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 2016)

Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay D.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shavchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S., Krzywicki M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J. and Marra M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

PUBMED

2 (bases 1 to 2016)

Strausberg R.

Direct Submission

Submitted (22-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

Contact: amadana@systemsbiology.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Place: 84 Row: 0 Column: 9.

Location/Qualifiers

1..2016

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:49020 IMAGE:6052140"

/tissue_type="Testis, embryonal carcinoma"

/clone_lib="NIH_MGC_92"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..2016

/gene="ERG"

/db_xref="LocusID:2078"

/db_xref="MIM:165080"

104..1543

/codon_start=1

/product="v-ets erythroblastosis virus E26 oncogene like, isoform 1"

/protein_id="AAH40168.1"

/db_xref="GI:25304066"

/translation="MASTIKKALSVSDQSLFECAYGTPHLAKTMTASSSDYQQT SRMSPRVQQDLWSQPPARVTIKMECPNSQVNSRNSPDECSVAGKXVGSFDTVGM NYGSYMEKHMPPNMTNERRVIVPADPTLMTSDHVRQWLEWAVKYLGPDPVNLFF QNIDKELCKMTKDDFQRLTYSYNADILLSHLYLRETPHLTSDVDVKALQNSPRL MHWNTGGAARFEPNTSVPEATORITRPDLVPBPPRRSAWTHGHPTQSKAAQPS PSTVPTKDDQRLDPYAILQPTSSRLANRSGGQIQWFLLELLSDSSNSCITWEG TNGEFKMDPDEVARWGRKSKPNNTDKLSRALRYYYDKNIMTKVGRKYAKFDF HGIAQALQPHPPESLYKPSDLPYNGSTHAHPQGNFVAPHPALPVTSSSFFAAPN PYWNSPTGGTYENTRLTPSHMPSHLGTY"

446..694

/note="SAM_PNT; Region: Sterile alpha motif (SAM)/Pointed domain"

/db_xref="CD:pfam02198"

1031..1282

/note="Ets; Region: Ets-domain"

/db_xref="CD:pfam00178"

ORIGIN

Alignment Scores:

Pred. No.: 4,75e-153 Length: 2016

Score: 2313.00 Matches: 429

Percent Similarity: 91.44% Conservative: 9

Best Local Similarity: 89.56% Mismatches: 13

Query Match: 94.64% Indels: 28

DB: 9 Gaps: 2

US-09-902-772-2 (1-451) x BC040168 (1-2016)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20

Db 104 ATGCCAGCACTATTAAAGGAAGCTTATCAGTTGTAGTGAGGACCACTGCTGTGTAG 163

Qy 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40

Db 164 TGTGCTTACGACGACCACTGCTGCTTACAGAGATGACCGCTCTCTCCACCGAC 223

Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60

Db 224 TATGGACAGACTTCCAAAGATGAGCCACGCTCCCTCAGCAGGATGGTGTCTCAACCC 283

Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80

Db 284 CCAGCCAGGGTCCACCATCAAAATGGAAATGTAACCTTAGCCAGGTGAATGCTCAAGAAC 343

Qy 81 SerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnVal 100

Db 344 TCTCCTGATGAATGCAGTGTGGCCAAAGCGGGAAGATGTTGGCGCACACACCGCTT 403


```
QY 42 GlyGlnThrSerLysMetSerProArgValProGlnAspTrpLeuSerGlnProPro 61
Db 292 GGCAGAGATCCAGATGAGTCCAGAGTCCCTCAGCAGGACTGGCTGTCTCAAGCCCA 351
QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db 352 GCCAGGGTCCACATCAAGATGGAGTCCAACTGATCAGGTGAATGGTTCAGGAATCA 411
QY 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGly 101
Db 412 CCTGATGAGTCAGTGTGAACARAGGTGGGAAGATGGTGGCAGCCGGATACCTGTGGG 471
QY 102 MetAsnTrpGlySerTrpMetGluGlnLysHisIleProProProAsnMetThrAsn 121
Db 472 ATGACTACGGCAGCTACATGAGGAGAGATGTCGCCCTCCCAATATGACACAAAT 531
QY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
Db 532 GAGCCAGAGTATCGTCCCTGCAGATCTTACTCTGTGGACACAGACCATGTCGGACAG 591
QY 142 TrpLeuGluTrpAlaValLysGluTrpGlyLeuProAspValAspIleLeuLeuPheGln 161
Db 592 TGGCTGGAGTGGCGGTGAAGAATATGCGCTCCTCGATGTGGACGTCTTACTATTTCAG 651
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
Db 652 AATATCGATGGAGAGAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGGCTCACGCCG 711
QY 182 SerTrpAsnAlaAspIleLeuLeuSerHisLeuHisTrpLeuArgGluArg 198
Db 712 AGCTACATGCCAGCATCTCTCTCACATCTCCACTACCTCAGAGAGACTCCCTTCCA 771
QY 198 ----- 198
Db 772 CATCTGACTTCGATGAGCTTGATAAGCTTTACAAACTCTCCACGCTTAATCATGCC 831
QY 199 ----- GlyAlaThrPheIlePheProAsnThrSerValTrpProGluAlaThr 214
Db 832 AGAAACACAGGGGTGCAGCTTTATTTTCCCAATACTTCAGTATATCCCGAGAGCTACG 891
QY 215 GlnArgIleThrArgProAspLeuProTrpGluGlnAlaArgArgSerAlaTrpThr 234
Db 892 CAAGAATATCAACTAGGCCAGATTTACCTTATGAGCTCCAGGAGATCAGCGCTGGACC 951
QY 235 SerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrValPro 253
Db 952 GGCACAGCCACTCACCCCTCAGTCCAAAGCTGCTCAGCCATCTCCCTCTGCAGTGCCC 1011
QY 254 LysThrGluAspGlnArgProGlnLeuAspProTrpGlnIleLeuGlyProThrSerSer 273
Db 1012 AAAACTGAAGACCCAGCGCTCCTCAGTTAGATCCTTACCAGATCCTGGGACCGACAGTAGC 1071
QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeu 293
Db 1072 CCCTTGCTTAATCCAGGTAGTGGCAGATCCAGCTGTGGCAGTCTCTCTCGAACTCCTG 1131
QY 294 SerAspSerSerAspAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
Db 1132 TCAGACAGCTCCAACTCCACTGTCATCCTCTGGGAGGACCAACAGCGGAGTTCAGAGT 1191
QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333
Db 1192 ACAGACCCGGAGAGTGGCTCGCGCTCGGGGAGAGAGAGCAAGCCCAACATGAAC 1251
QY 334 TyrAspLysLeuSerArgAlaLeuArgTrpTrpTrpAspLysAsnIleMetThrLysVal 353
Db 1252 TATGACAGCTCAGCGCGCCCTCCGCTACTACTACTAGCACAAACATCATGACCCAGGTG 1311
QY 354 HisGlyLysArgTrpAlaTrpLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
Db 1312 CACGGGAAGCGCTACGCTTACAGTTGACTTCCACGGGATTCGCCAGCGCTGCAGGCC 1371
```

```
QY 374 HisProProGluSerSerMetTyrLysTrpProSerAspLeuProTrpMetSerSerTrp 393
Db 1372 CACCCTCTCGAGTCGTCCTGTACAGTACCCCTCCGACTCGCATACATGGGCTCTAT 1431
QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413
Db 1432 CAGCCACCCACAGAGATGACATTTGTGTCTCCACCCCTCCGCTCTCCAGTCACA 1491
QY 414 SerSerSerPhePheAlaAlaProAsnProTrpTrpAsnSerProThrGlyGlyIleTrp 433
Db 1492 TCTTCCAGTTCTTCTTCTTCCCGAACCATATGGAATTCACCGACTGGGGCATCTAC 1551
QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTrpTrp 451
Db 1552 CCGAACACTAGGTCCTCCAGCCAGCATATGCCCTCTCACCCTGGGACCTACTAC 1605

RESULT 8
XLAJ4126 2012 bp mRNA linear VRT 21-JAN-2000
LOCUS Xenopus laevis erg gene (erg_E).
ACCESSION AJ224126
VERSION AJ224126.1 GI:5420047
KEYWORDS ERG gene; transcription factor.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 Baltzinger,M., Mager-Heckel,A.M. and Remy,P.
AUTHORS X1 erg: expression pattern and overexpression during development
TITLE plead for a role in endothelial cell differentiation
JOURNAL Dev. Dyn. 216 (4-5), 420-433 (1999)
MEDLINE 20093678
PUBMED 10633861
REFERENCE
2 (bases 1 to 1012)
AUTHORS Baltzinger,M.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1998) Baltzinger M., UPR9005, MIMD, Centre
National de la Recherche Scientifique, 15, RUE Rene Descartes,
67084, FRANCE
FEATURES
source
1..2012
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="lambda ZAPII-erg E"
/clone_lib="lambda ZAPII"
/dev_stage="stage 30 embryo"
1..2012
/gene="erg"
25..1595
/gene="erg"
/function="transcription factor"
/codon_start=1
/protein_id="CAB46567.1"
/db_xref="GI:5420048"
/db_xref="GOA:Q9W629"
/db_xref="SPTREMBL:Q9W629"
/translation="MIOCTVDPSSAHIKEALSVVSEDSQSLFECTYGTGPHLTKTEMTASS
SDYGTGSKMSPRVPOQDLNLSQPPSVTTIKMCNPSQVNGSRSSPDCCGIGKSGKMG
GSDNVMNGSVTEEXHISPPNMTNERSVIYPADPTLWSTDHVRQWLKWAKEYGLP
DVVLFQNIIDGKELCKMTKEDFHLTPSYNADILLSHLYLERGASRIFPNSSVYQ
DANQRIPSKQDLSYSPRSAMTHPPSKASQPSSTTVFKEDPRQDLPQILGPT
SSLANPGSQIQWLQFLLLELSDSNCSITWEGTNGBFKMTDPDEARRWGERSK
PNMYDKLGRALRYYYDKNIMTKHGKRYAKYKDFHGIQAQLOPHPESTMYKPSBL
PYMNSYHAHPQKNFVAPHPALPVTSSFFFAAPNAYMNSPTGSIYENTRLPASHMSS
HLGTYY"
ORIGIN
Alignment Scores: 1.46e-144 Length: 2012
Pred. No.: 2192.00 Matches: 402
```

```

Percent Similarity: 94.63%      Conservative: 21
Best Local Similarity: 89.93%    Mismatches: 22
Query Match: 89.69%             Indels: 2
DB: 5                            Gaps: 2

US-09-902-772-2 (1-451) x XLAJ4126 (1-2012)

QY 5 IleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCysAlaTyrGly 24
DB 258 ATCAAGAAGCTCTTTTCAGTGGTGAAGATCAGTCTGTTATTTCGAGTGCACTACGGA 317
QY 25 SerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerSerGluTyrGlyGlnThr 44
DB 318 AGCCACATCTTACTAAACGGAGATGACCGCATCTTCTCCAGTGACTATGGCAACC 377
QY 45 SerLysMetSerProArgValProGlnGlnAspThrLeuSerGlnProProAlaArgVal 64
DB 378 TCGAAATGAGTCCACGTCGTGCTCAACAAGACTGGCTCTCACAGCTCCATCCAGGGTG 437
QY 65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84
DB 438 ACCATCAAGATGAGTATGACACCTAGTCAATGATCAAGGAGCTCCCGGATGAC 497
QY 85 CysSerValAlaLysGlyGlyLysMetValSerSerSerSerSerSerValGlyMetAsnTyr 104
DB 498 TGCAGTATAGGAAAGGAAGCAAAATGGGTGGAGGTCGGATATATGTTGGAATGAAATAT 557
QY 105 GlySerTyrMetGluGluLysIleProProProAsnMetThrThrAsnGluArgArg 124
DB 558 GGCAGCTACATAGAAAGAAACACATTTCTCCCCAAACATGACAAACATGAGAGAGA 617
QY 125 ValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGlu 144
DB 618 GTGATTGTACCTGCTGACCTACTTGTGGAGCACAGATCATGTCCAGACAGTGGCTAGAG 677
QY 145 TrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLysPheGlnAsnIleAsp 164
DB 678 TGGGCAATAAAGATACCGCTTCCAGATGTCGATGTCTCTCTTCCAAACATCGAC 737
QY 165 GlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsn 184
DB 738 GCGAGGANTATGCAAGATGACCAAGAGAGATTTCCATCGACTCACGCCAAGCTTACAT 797
QY 185 AlaAspIleLeuSerHisIleuHisTyrLeuArgGluArgGlyAlaThrPheIlePhe 204
DB 798 GCTGATATCTCTCTCGCATCTTCACTACCTCAGAGAAAGAGTGGCTCTTTTATTTTC 857
QY 205 ProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgProAspLeuPro 224
DB 858 CCAAACTCATCGTTTATCAAGCGCAACCAAGATCCCGAGCAGCAGATTTATCA 917
QY 225 TyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThrGlnSerLysAla 244
DB 918 TATGAGCCCTCAAGGAGATCTGCTTGGACAAATCATCTCTGCACCA---CCATCAAAAGCA 974
QY 245 ThrGlnProSerSerThrValProLysThrGluAspGlnArgProGlnLeuAspPro 264
DB 975 TCTCAACCA---TCTACCAACAGTTTCCAAACAGAAAGCCCGAGCCACAGCTAGATCTCT 1031
QY 265 TyrGlnIleLeuGlyProThrSerArgLeuAlaAsnProGlySerGlyGlnIleGln 284
DB 1032 TATCAGATCTTGGGCCAACAGCAGCCGACTTGCAAATCCAGGAAGTGGCGAGATTCCAG 1091
QY 285 LeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnSerAsnSerAsnSer 304
DB 1092 CTCGCAATCTCTTACTGAACTGCTTTTCGATAGTTCATCTCCAACTCAACTGCAATCTACTGG 1151
QY 305 GluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArgTrpGly 324
DB 1152 GAAGGAACCAATGGAGAAATTAAGATGACCGGATCTCTGATGAAGTGGCAGACGTTGGGG 1211
QY 325 GluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyr 344

```

```

1212 GAGAGGAAAGCAACCAACCAACATGAACTATGATAAGCTCAGCGCTGCCTTACTTACTAC 1271
345 TyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAspPhe 364
1272 TAGCATAAATAATATGATGACTAAAGTCCATGCAAGCGCTATGCTTACAAATTTGATTTT 1331
365 HisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMetTyrLysTyrPro 384
1332 CATGGGATTTGCTCAGCTCTTCAACCTCATCTCCAGATCAACTATGTACAAGTATCCC 1391
385 SerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetAsnPheValAla 404
1392 TCAGAACTTCCATATATGAGCTCATACCATGACATCCACAGAAGATGAACCTTTGTGTC 1451
405 ProHisProProAlaLeuProValThrSerSerSerPhePheAlaAlaProAsnProTyr 424
1452 CCCATCCCGCAGCTTACCTGTGACATGCTCAAGTTCTTTGCGACCCCTAATGCATAC 1511
425 TrpAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAlaAlaHisMetPro 444
1512 TGGAAATTCACCAACTGGAAGTATTTATCCAAATACTCGGCTGCCAGCTAGCCATATGCT 1571
445 SerHisLeuGlyThrTyrTyr 451
1572 TCTCATCTTGGAACTTACTAT 1592

RESULT 9
HUMERG2
LOCUS HUMERG2 3166 bp ss-RNA linear PRI 08-NOV-1994
DEFINITION Human erg2 gene encoding erg2 protein, complete cds.
ACCESSION MI7254.1
VERSION MI7254.1 GI:182186
KEYWORDS erg 2 protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3166)
AUTHORS Rao, V.N., Papas, T.S. and Reddy, B.S.
TITLE erg, a human ets-related gene on chromosome 21: alternative
JOURNAL Science 237 (4815), 635-639 (1987)
MEDLINE 87263429
PUBMED 3299708
COMMENT Original source text: Human cell line COLO 320, cDNA to mRNA, clone
lambda 12.
FEATURES
source 1..3166
/organism="Homo sapiens"
/moi_type="genomic RNA"
/db_xref="taxon:9606"
/map="21q22.3"
257..1645
/gene="ERG"
257..1645
/gene="ERG"
/notes="erg 2 protein"
/codon_start=1
/protein_id="AA52398.1"
/db_xref="GI:182187"
/db_xref="GDB:G00-119-884"
/translation="MTQVDPAAHFKELSVSVSEDSLSFECAGYGPHLAKTENTASS
SDYGTSKMSPRVPOQWLSQPPARTIMECNPSQVNSRNSPDECSVAKGKMGV
SPDTGVNYSYMEKEKMPNNMTTNERAVIVPADPTLWSTDRQWLWVAVKEYGLP
DVNILLFQNTDGKELCKMTKDFQRLTPSYNADILLSHLYLFEETPLHLTSDVDKA
LQNSPLRMHANTDLPVEPRRSATWGHTGHPTQSKAAQPSPTVPKTEQDRQLDPY
OILGPTSSRLNPGSGQIOLWQFLLELLSDSSNSCTTWEETNGEFKMTDPDEVARRW
GERKSPNNYDKLSALARYIDKNITMTKVHGKRYAKPDFHGIQALQPHPESSLY
KYPSDUPYMGSYHAHPQKNFVAPHPFALPVTSSSFFAAPNPYNSPTGGLYFNTRLP
TSHMPGHLGTY"
repeat_region 291..298
/notes="8-bp repeat sequence"
repeat_region 509..516

```

/note="g-bp repeat sequence"
3143..3149
polya_signal 227 bp upstream of PvuII site; chromosome 21.
ORIGIN
Alignment Scores:
Pred. No.: 1.25e-143 Length: 3166
Score: 2182.00 Matches: 407
Percent Similarity: 91.72% Conservative: 14
Best Local Similarity: 88.67% Mismatches: 24
Query Match: 89.28% Indels: 14
DB: 9 Gaps: 4
US-09-902-772-2 (1-451) x HUMERG2 (1-3166)
QY 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
DB 281 GCAGCTCATATCAAGAGCCCTTATCAGTTGTGAGTGGACCAAGTCTGTTTGTAGTGT 340
QY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyr 41
DB 341 GCCTACGGAAGCCACACCTGGCTAAGACAGAGATGACCGGTCTCTCTCCAGGACTAT 400
QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61
DB 401 GGACAGACTTCAAGATGAGCCCGCCGCTCCCTCAGCAGGATTGGCTCTCAACCCCCCA 460
QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
DB 461 GCCAGGGTCACCATCAAAATGGATGTAAACCTAGCCAGGTGAATGGCTCAAGGAATCT 520
QY 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGly 101
DB 521 CCGTGAATGAGTCAGTGTGGCCAAAGCGCGGAGATGGTGGCGACGCCACACCGTTGGG 580
QY 102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsn 121
DB 581 ATGAATCTACGGCAGTACATGGAGAGAGAGACATGCGCCACCCCAACATGACCGAGAC 640
QY 122 GluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141
DB 641 GAGCGCAGAGTTATCGTGCCAGCAGATCTACGCTATGGAGTACAGACCATGTGCGGCG 700
QY 142 TrpLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
DB 701 TGCGTGGAGTGGCGGTGAAAGNATATGGCTTCCAGACGTCAACATCTGTTATTCCAG 760
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB 761 AACATCGATGGAGAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCCGC 820
QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
DB 821 AGCTACACGCGCAGATCCTCTCTCAGATCTCCACTACCTCAGAGAGACTCCCTCTTCCA 880
QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
DB 881 CATTGG-----ACITCA-----GATGATGTTGATAAGCGCTTACAAACTCTCCA 925
QY 222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTyr 233
DB 926 CGGTTAATGCATGCTAGAAACACAGATTACCATATGAGCCCCCGCCAGAGATCAGCCTGG 985
QY 234 ThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrVal 252
DB 986 ACCGTCACGGCCACCCAGCCCGCCAGTGCAGAAAGTGTCTCAACCATCTCTCTCCACAGTG 1045
QY 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272
DB 1046 CCCAAACTGAAGACGAGCGTCTCAGTAGATCTTATCAGATTCCTTGGACCAACAGT 1105
QY 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeu 292

1106 AGCCGCTTGCRAATCCAGGCAGTGGCCAGATCCAGCTTTGGCAGTTCCTCTGGAGCTC 1165
QY 293 LeuSerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLys 312
DB 1166 CTGTGGAGAGCTCCCAACTCCAGCTCATCAGCTGGGAAGGACCAACCGGGAGTTCAAG 1225
QY 313 MetThrAspProAspGluValAlaArgAArgTyrGlyLysSerLysProAsnMet 332
DB 1226 ATGACGGATCCCGACGAGGTGGCCGCGCTGGGAGAGCGGAGAGCAACCAACATG 1285
QY 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLys 352
DB 1286 AACTACGATAAGCTCAGCCGCGCCCTCCGTACTACTATGCAAGAACATCATGACCAAG 1345
QY 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
DB 1346 GTCCATGGAGAGCGTACGCTACAGTTCCGATTCCTCCAGGATCGCCAGGCCCTCCAG 1405
QY 373 ProHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
DB 1406 CCCACCCCGGAGTCACTCTGTACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCC 1465
QY 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProVal 412
DB 1466 TATCAGCCCGCCACAGAGATGAATTTGTGGCGCCGCCCTCCAGCCCTCCCGGTG 1525
QY 413 ThrSerSerSerPhePheAlaProAsnProTyrTyrAsnSerProThrGlyGlyIle 432
DB 1526 ACATCTTCAGTTTGTGTCGCCAAACCCCATCTGGAATTCACCAACTGGGGGTATA 1585
QY 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
DB 1586 TACCCCAACACTAGGCTCCCGACCAAGCATATGCTTCTCATCTGGGCACTTACTAC 1642
RESULT 10
LOCUS AB031088 1808 bp mRNA linear ROD 07-AUG-2001
DEFINITION Rattus norvegicus VESP14 mRNA for vascular endothelial cell specific protein 14, complete cds.
ACCESSION AB031088
VERSION AB031088.1 GI:15128488
KEYWORDS vascular endothelial cell specific protein 14.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1808)
AUTHORS Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and Komurasaki,T.
TITLE Identification of VESP14,a vascular endothelial cell specific protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1808)
AUTHORS Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and Komurasaki,T.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1999) Tomomi Aoki, Taisho Pharmaceutical Co., Ltd., Medicinal Research Laboratories, Molecular Biology Lab., 1-403, Yoshino-cho, Omiya, Saitama 3308530, Japan (E-mail:sl17704@ccm.taisho.co.jp, Tel:048-663-1111, Fax:048-652-7254)
FEATURES
source 1..1808
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/tissue_type="liver"
gene 1..1808
/genes="VESP14"
CDS 63..1430
/genes="VESP14"
/note="similar to human erg1"

/codon_start=1
/product="vascular endothelial cell specific protein 14"
/protein_id="BAB62744.1"
/db_xref="GI:15128489"
/translation="MASTIKKALSVWSKQSLFECAYGTPLHAKTEMTSSSDYGT
SKMSPRVQDMLSPARVTIKMECPNSOVNGSRNSPDECSVTGKGMVGSPTDVG
SYGSMGEKMPNPNTNRPVLPADPTLMSDTHVRQWLEWAVEYGLDVLVLLF
QNIQKELCKMTKDDFORLPSYNADILLSHLYLRETPLPLHLSDDVDKALQNSRL
MHRANTDLPFPPRSWTGSHPTPOSKAAQSPSPFKTEDEQFOLDPIYQIGTIS
SLANPQSGQQLQWQFLELLSSDSSNCSITWETNGTEFNTDPEDEVRWGRKSKP
NMNYDKLSRALRYFDKNIWTKVGRYAYKDFPHGIAQALPHPESSLYKYPSPDP
YMGSHYTHPQKQNFVAPHPFPALPVTSSFFATNPYNPSPTGGIYENTRLPASHMPSH
LGTYY"

ORIGIN

Alignment Scores:
Pred. No.: 1,056-143 Length: 1808
Score: 2179.00 Matches: 406
Percent Similarity: 91.30% Conservative: 14
Percent Local Similarity: 88.26% Mismatches: 26
Query Match: 89.16% Indels: 14
DB: 10 Gaps: 4

US-09-902-772-2 (1-451) x AB031088 (1-1808)

QY	1	MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu	20
DB	63	ATGGCCAGCAGCTATTAAAGAGGCGCTTGTCAGTTGTGAGCAGGACCACTCACTATTGAG	122
QY	21	CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu	40
DB	123	TGTGCTACGGAACCCACACTGCTGCTAAGACAGAAATGACCGCATCTCTTCAGTGAC	182
QY	41	TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro	60
DB	193	TATGCCAGACATCCAGATGATGCTCCAGATGCCACACAGAGTGGCTGTCTCAGCCC	242
QY	61	ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn	80
DB	243	CCAGCCAGGCTCACCATCAAGATGAGTGCACACCTAGCCAGGTGAACGGTTCAGGAAC	302
QY	81	SerProAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal	100
DB	303	TCACTGACGAATGACGTGTGACCAAGTGGGAGATGGTGGCAGCCCTGATCTGTG	362
QY	101	GlyMetAsnTyrGlySerTyrMetGluGlyLysHisIleProProProAsnMetThr	120
DB	363	GGAAATGAGCTACGGCAGCTACATGAGGAGGAGACATGCGCGCCCAATATGACCACG	422
QY	121	AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg	140
DB	423	AATGAACGAGAGTATGCTGCTCCAGATCTCTCTGTGAGCAGACATGTCGG	482
QY	141	GlnTyrLeuGluTyrAlaValLysGlyTyrGlyLeuProAspValAspIleLeuPhe	160
DB	483	CAGTGGCTGAGTGGGAGTGAAGAATAGCGCTCTTAGAGCTGGAGCTCTTATTATT	542
QY	161	GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr	180
DB	543	CAGAATATTGACGGGAGGAGTGTGCAAGATGACAAAGATGACTTCCAGGCGCTCACT	602
QY	181	ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAla	200
DB	603	CCGAGCTACAAATGCCGACATTTCTCTCACTCTCCACTCTCCACTCAGAGACTCCCTT	662
QY	201	ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg	220
DB	663	CCACATCTG-----ACATCC-----GACGACGTTGTAAGCGCTTTACAAACTCT	707
QY	221	Pro-----AspLeuProTyrGluGlnAlaArgSerAla	232
DB	708	CCACGGTTAATGATCTAGAAACACAGATTTTACCTTATGAGCTCCCGAGGATCAACC	767

QY	233	TyrThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThr	251
DB	768	TGAGCGGGCCACAGCAGCCACCCCTCAGTCCAAAGCTGCCAGCAATCTCCCTCCACA	827
QY	252	ValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThr	271
DB	828	GTGCCCAAACTGAAGACCAGCGTCTCAGTTAGATCTTACCAATCTCTGGGACCCACC	887
QY	272	SerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuGlu	291
DB	888	AGTAGTCGCTTGTCTAAATCCAGTAGTGCACAGATCCAGCTGTGCAGTTCCTGTAGNA	947
QY	292	LeuLeuSerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGlyLysPhe	311
DB	948	CTCCTGTCTGACAGCTCCAACTCCAACTGATCATCTGGGAAGGACCAACGGGAGTTC	1007
QY	312	LysMetThrAspProAspGluValAlaArgTyrGlyGluArgLysSerLysProAsn	331
DB	1008	AAGATGACAGACCCGAGTAGGTGGCTCGGCCCTGGGGAGCGGAGGAGGAGCCCAAC	1067
QY	332	MetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThr	351
DB	1068	ATGAACTATGACAACTCAGCGCTGCTCCCTCTACTACTAGACAGAAAACATCATGACC	1127
QY	352	LysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeu	371
DB	1128	AAGGTGCACGGGAAGCGCTATGCTTACAGTTTGATTTCCACGGGATTCGCCAGGCCCTG	1187
QY	372	GlnProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSer	391
DB	1188	CAGCCCACTCCCGCGAGTCTGCTGTACAGTACCTCCCTCCGCTGCGTACATGGGC	1247
QY	392	SerTyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuPro	411
DB	1248	TCTATCACACACCCCGCAGAGATGAACTTTGGGTCTCCCGCTCCCGCCCTCCCA	1307
QY	412	ValThrSerSerSerPhePheAlaProAsnProTyrTyrAsnSerProThrGlyGly	431
DB	1308	GTCACTCTTCAGTTTCTTGTCTACCCCGACACCATCTGGAATTCGCGGATGGGGC	1367
QY	432	IleTyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr	451
DB	1368	ATCTACCCGAACAGTAGGCTCCGAGCAGCATATGCTCTCTCAGCTGGGACCTACTAC	1427

RESULT 11

AB073079 2133 bp mRNA linear ROD 17-OCT-2001
Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,
transcript variant 2, complete cds.

AB073079
AB073079.1 GI:16197542

Mus musculus (house mouse)
Mus musculus

REFERENCE
1 Ozawa, R., Noguchi, H., Taylor, T.D., Takeda, T., Hattori, M. and Sakaki, Y.

Mus musculus Erg mRNA
Published Only in Database (2001)

2 (bases 1 to 2133)
Ozawa, R., Noguchi, H., Taylor, T.D., Takeda, T., Hattori, M. and Sakaki, Y.

Direct Submission
Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

Location/Qualifiers
1. .2133
/organism="Mus musculus"

FEATURES

source

JOURNAL Dev. Dyn. 216 (4-5), 420-433 (1999)
 MEDLINE 20099678
 PUBMED 10633861
 REFERENCE 2 (bases 1 to 1798)
 AUTHORS Baltzinger, M.
 TITLE Direct Submission
 JOURNAL Submitted (24-FEB-1998) Baltzinger M., UPR9005, MDCD, Centre
 National de la Recherche Scientifique, 15, RUE Rene Descartes,
 67084, FRANCE
 FEATURES Location/Qualifiers
 source 1..1798
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="lambda ZAPII-erg A"
 /clone_lib="lambda ZAPII"
 /dev stage="stage 30 embryo"
 1..1798
 /gene="erg"
 /gene="erg"
 321..1778
 /function="transcription factor"
 /codon start=1
 /protein id="CAB46566.1"
 /db_xref="GI:5420046"
 /db_xref="GOA:Q9W700"
 /db_xref="SPRAMEL:Q9W700"
 /translation="MQMAFIPTDYDYSNALSVSVDQSLEFCTVGTPLTKTEMTA
 SSSDYQTKMSPRVQDWSLPSPRTVTKMCPNSQNGSRSSPDSCIGKSKM
 GGSNDVMGYVIEBKHSIPNMTNERRVVPADPTLWSTDHVRLQWLWAIKEYG
 LPDVLFPNIDGKELCKMKEDFHLTPSYNADILLHLHYLRTPPLHLSDDVD
 KALQNSPLMARHTGASFI PNSSVYODANORIPSRDLSYEPERRSAWTHNPAPP
 SKASQSTVTPKTEDEPDLPYQILGPTSSRLANPGSGOIQWQLLELLSDSNSN
 CIWEGINGEFKMTDEVARNGERKSKENMYDKLSRALRYVYDKNTWTKVGRY
 AYKDPFGIAQALPDPESTMYKIPSELPYMSYHAFQPMFVAPHPFPALPVTSSS
 FFAAPNAYMNSPTGSIYPTRLPASHMSHLGTY"
 ORIGIN

Alignment Scores:
 Pred. No.: 6, 48-142 Length: 1798
 Score: 2153.50 Matches: 399
 Percent Similarity: 88.98% Conservative: 21
 Best Local Similarity: 84.53% Mismatches: 23
 Query Match: 88.11% Indels: 29
 DB: 5 Gaps: 3
 US-09-902-772-2 (1-451) x XLAJ4125 (1-1798)
 Qy 7 GluAlaLeuSerValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26
 Db 366 GAAGCTCTTTTCAGTGTGAGTGAAGATCAGTCGTATTTCAGTGCACCTACGGAACGCCA 425
 Qy 27 HisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyrGlyGlnThrSerLys 46
 Db 426 CATCTTACTAAACGGAGATGACCGCATCTTCTCCAGTGCATATGGGCAAACTCGAAA 485
 Qy 47 MetSerProArgValProGlnGlnAspTrpLeuSerGlnProAlaArgValThrIle 66
 Db 486 ATGAGTCACCTGTGCTCAACAAGATCGGTCTTCACAGCTCCATCCAGGGTGACCATC 545
 Qy 67 LysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAspCysSer 86
 Db 546 AAGATGGAATCAACCTTCTCCAGTCAATGATCAAGGAGCTCCCGGATGCTGCAGT 605
 Qy 87 ValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySer 106
 Db 606 ATAGAAAAGGAAGCAAAATGGGTGGAGGTGCGGATATGTTGGAATCAATATTGGCAGC 665
 Qy 107 TyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgArgValIle 126
 Db 666 TACATAGAGAAACACATTTCTCCCCCAACATGACCAACATGAGAGAGAGTGATT 725
 Qy 127 ValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAla 146

Db 726 GTACCTGCTGACCCCTACTTTTGTGGAGCACAGATCATGTCAACAGATGGCTAGATGGCA 785
 Qy 147 VallysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys 166
 Db 786 ATAAAGATACGGCTTCCAGATGTGATGTGCTCTCTGTTTCCAAACATCAGCGCAAG 845
 Qy 167 GluLeuCysLysMetThrLysAspPheGlnArgLeuThrProSerTyrAsnAlaAsp 186
 Db 846 GAATTTATGCAAGATGACCAAGAAGATTTCATGCTGACTCAGCGCAAGCTACAATGCTGAT 905
 Qy 187 IleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
 Db 906 ATCCCTTCTCGCATCTTCACTTACTCTCAGAGAAACTCCTCTTCCCATTTGACTTCAGAT 965
 Qy 199 -----Gly 199
 Db 966 GATGTTGATAAGGCGCTTACAAAACCTCCACGGTAAATGTCATGTAGAAATACAGGAGGT 1025
 Qy 200 AlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 219
 Db 1026 GCCTCTTTTATTTTCCCAACTCATCCGTTTATCAGAGGCAACCAAGATCCCCAGC 1085
 Qy 220 ArgProAspLeuProTyrGluGlnAlaArgSerAlaTyrThrSerHisSerHisPro 239
 Db 1086 AGGCAAGATTTATCATATGAGCCCTCAGAGAGATCTGCTTGGCAAAATCATCTCGACCA 1145
 Qy 240 ThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArg 259
 Db 1146 ---CCATCAAAAGCATCTCAACCA---TCTACCAAGTTCCTCAAAACAGAGAGCCCCAGG 1199
 Qy 260 ProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGly 279
 Db 1200 CCACAGCTAGATCTCTTATCAGATCTTGGGCAACCAAGCAGCGCATTTGCAAAATCCAGGA 1259
 Qy 280 SerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuSerSerSerSerSerSer 299
 Db 1260 AGTGGGAGATTCAGCTCTGCAATCTTACTGGAACCTGCTTTCGATAGTTCCTCACTCC 1319
 Qy 300 AsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluVal 319
 Db 1320 AACTGCATTACTTGGGAAGGAACCAATGGAGAAATTTAAGATGACCGATCCCGATGAAGTG 1379
 Qy 320 AlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArg 339
 Db 1380 GCAAGACGTTGGGGGGAGAGAAAGCAACCCCAACATGAACTATGCAAGCTACGCCGT 1439
 Qy 340 AlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAla 359
 Db 1440 GCACCTTCTGTTACTACTACGATAAAATATTTATGACTAAGTCCATGGCAAGCGCTATGCT 1499
 Qy 360 TyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerSer 379
 Db 1500 TACAAATTTGATTTTTCATGGGATTCCTCAGGCTCTTCAACCTCATCTCCAGATCAACCC 1559
 Qy 380 MetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLys 399
 Db 1560 ATGTACAAGATATCCCTCAGAACTTCATATATGACTCATACCATGCAATCCACAGAG 1619
 Qy 400 MetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAla 419
 Db 1620 ATGAACCTTTGTTGCCCCCATCCCGCAGCTTTACCTGTGACATCGTCAAGTTCTTTTGA 1679
 Qy 420 AlaProAsnProTyrTyrAsnSerProThrGlyIleTyrProAsnThrArgLeuPro 439
 Db 1680 GCCCTTAATGCAATCTGGAATTCACCAACTGGAAGTATTTATCCAAATACTCCGCTGCCA 1739
 Qy 440 AlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1740 GCTAGCCATATGCTCTCTCATCTTGGAACTTACTAT 1775

RESULT 13
 AB073078

LOCUS AB073078 2131 bp mRNA linear ROD 17-OCT-2001
 DEFINITION Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,
 transcript variant 1, complete cds.
 ACCSSION AB073078
 VERSION 1
 KEYWORDS AB073078.1 GI:16191716
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.
 AUTHORS Mus musculus Erg mRNA
 TITLE Published Only in Database (2001)
 JOURNAL 2 (bases 1 to 2131)
 REFERENCE Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.
 AUTHORS Direct Submission
 TITLE Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical
 JOURNAL and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 FEATURES
 source
 1. 2131
 /location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /chromosome="16"
 /clone="Erg-1"
 1. 2131
 /gene="Erg"
 148..1536
 /gene="Erg"
 /note="homolog of human ets-related gene (ERG), transcript
 variant 1"
 /codon_start=1
 /protein_id="BAB69948.1"
 /db_xref="GI:16191717"
 /translation="MIGTVDPDPAHKEALSVVSEPSQSLPECAYGTFPHLAKTMTASS
 SSDYGTQSKMSPRVQDDWLSQAPARVITKMCNPSQVNGSRNPSPECVNGKGNV
 PFDVGVSGYSYVEKEKVPFPNMTNERRVIVPADPLMTSDHVRQWLWAVKEVGLL
 DVDFLLQNDIGELCKMTKDDFORLTPS VNADILLSHLYRETPLPLTSDVDVKA
 LQNSPLMHARNLDLPEPRRSANWGHSHLTQSKKQAPSPSAVPEKTDQRQLDPY
 QIGPTSRRLANPGSGIQLWQFLELLSDNSNCLTWEGTNGERFAMDPDFEVARRW
 GERKSPNMYDKLSRLRYYYDKNIMTKVHGKRYAYKDFHGIAGLQHPPEPSLY
 KPSDLPMGYSYHAHPQKXNFVSPHPALPVTSSFPASPYPNPSFTGGIYENTRLP
 ASHMPSHLGYTY"

Alignment Scores:
 Pred. No.: 1.17e-141 Length: 2131
 Score: 2151.00 Matches: 402
 Percent Similarity: 91.07% Conservative: 16
 Best Local Similarity: 87.58% Mismatches: 27
 Query Match: 88.01% Indels: 14
 DB: 10 Gaps: 4
 US-09-902-772-2 (1-451) x AB073078 (1-2131)
 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
 172 GCAGCTCATATTAGGAGCCCTTGTAGTTGTGAGCGAGGACCACTTATTAGTGT 231
 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyr 41
 232 GCCTACGGAACGCCACACCTCGCTAAGACAGAGATGACCGCATCTCTTCCAGTACT 291
 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTTPLeuSerGlnProPro 61
 292 GCCACAGATCCAGATGATCCAGAGATCCCTCAGCAGGACTGGCTGTCTCAGGCCCA 351

ORIGIN
 Alignment Scores:
 Pred. No.: 1.17e-141 Length: 2131
 Score: 2151.00 Matches: 402
 Percent Similarity: 91.07% Conservative: 16
 Best Local Similarity: 87.58% Mismatches: 27
 Query Match: 88.01% Indels: 14
 DB: 10 Gaps: 4
 US-09-902-772-2 (1-451) x AB073078 (1-2131)
 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
 172 GCAGCTCATATTAGGAGCCCTTGTAGTTGTGAGCGAGGACCACTTATTAGTGT 231
 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyr 41
 232 GCCTACGGAACGCCACACCTCGCTAAGACAGAGATGACCGCATCTCTTCCAGTACT 291
 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTTPLeuSerGlnProPro 61
 292 GCCACAGATCCAGATGATCCAGAGATCCCTCAGCAGGACTGGCTGTCTCAGGCCCA 351

62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
 352 GCCAGGTCACCATCAGATGAGTGCACCTTAGTCAGTGAATGGTTCCAGGAACCTCA 411
 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGly 101
 412 CCTGATGAGTGCAGTGTGACAAAGGTGGAGATGTTGGGCGAGCCCGACTCTGGGG 471
 102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThrAsn 121
 472 ATGAGCTACGGCAGCTACATGAGGAGAGAGATGTCGCCCTCCCAATATGACCAAA 531
 122 GluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141
 532 GAGCGCAGAGATGATCGCTCGCATCTCTGTGGAGCAGACCATGTCTCGACAG 591
 142 TrpLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuPheGln 161
 592 TGGCTGGAGTGGCGCGTGAAGATATGCTCTCCGATGTGGAGCTTACTATTTCAG 651
 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
 652 AATATCGATGGAGAGAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGGCTCAGCG 711
 182 SerTyrAsnAlaAspIleLeuLeuSerHisIleHisTyrLeuArgGluArgGlyAlaThr 201
 712 AGCTACAATGCGCATCTTCTCTCATCTCCACTACCTCAGAGAGACTCCCTTCCA 771
 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
 772 CATCTG-----ACTTCC-----GATGACGTTGATAAGGCTTTACAAACTCTCCA 816
 222 -----AspLeuProTyrGluGlnAlaArgSerAlaTyr 233
 817 CGGTAAATGTCATGCCAGAAACACAGATTTACCTTATGAGCCTCCAGAGATCACCGTGG 876
 234 ThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrVal 252
 877 ACCGCCACAGCCACTCACCTCAGTCCAAAGTGTCTCAGCCATCTCCCTCTCAGTGTG 936
 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272
 937 CCAAAAACCTGAAGACAGCGCTCTCAGTTAGTCTTACCATGATCTGGAGCCGACCACT 996
 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeu 292
 997 AGCCGCTTGTCTAATCCAGGTAGTGGCAGATCCAGCTGTGGCAGTTCTCTGCTCGACTC 1056
 293 LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys 312
 1057 CTGTTCAGACAGCTCCCAACTCCACTCCATCACCTGGGAAGGCACCAACGGGGAGTTCAAG 1116
 313 MetThrAspProAspGluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMet 332
 1117 ATGACAGACCCGAGAGGTGCTCGCGCTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGATG 1176
 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 352
 1177 AACTATGACAAGCTCAGCGCGCCCTCCGCTACTACTACGACAAACATCATGACCAAG 1236
 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
 1237 GTGCACGGGAAGCGCTACCGCTACAGTTGTACTTCCAGGGATGTCAGGAGGAGGAGGAG 1296
 373 ProHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
 1297 CCCCACCTCTCTGAGTCTCGCTCCCTGTACAAGTACCCCTCCGACCTGCCATACATGGCTCC 1356
 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProVal 412
 1357 TATCAGCCCAACCCAGAGATGAATTTGTGTCTCCCCACCTCCCGCTCTCCAGTCTC 1416
 413 ThrSerSerSerPhePheAlaProAsnProTyrTyrTipAsnSerProThrGlyGlyIle 432

Db 1417 ACATCTTCAGTTCTTTGCTTCCCGAACCCATCTGGAATTCACCGACTGGGGGATC 1476

Qy 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyTyr 451

Db 1477 TACCCGAACACTAGGCTCCCGACGACCATATGCCCTCTCACCTGGGCACCTACTAC 1533

RESULT 14

AX657230 1291 bp DNA linear PAT 22-MAR-2003

LOCUS DEFINITION Sequence 3 from Patent WO03000724.

AX657230

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomati; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Luyten,F., de Bari,C. and Dell'Accio,F.

TITLE Polynucleotide sequences and vectors useful for the prevention or treatment of bone or cartilage-related disorders

JOURNAL Patent: WO 03000724-A 3 03-JAN-2003;

Tigenix N.V. (BE)

FEATURES

source

1. .1291

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

146..1228

note="unnamed protein product"

/codon_start=1

/protein_id="CAD79712.1"

/db_xref="GI:29160009"

/db_xref="RENTREMBL:CAD79712"

/translation="MGSPDTVMGVGSMYEEKMPPNMTNERRVIVPADPTLWST
DHVRQMLEAVKVGJFDPVNNILFQNDIKELCKMTKDDFORLTPPSINADILLSHUHY
LRRGAAFIPTNSVYPEATQRTITRPDLYPEPSRNTAGHPTPQSKAAQPSPT
EFTEDORQDLYQILGTPSSRLANPGSGQIQWQLFLLSDSSNSCITWEGTNG
VFKMTQDPEVARWGRKSPNNYDKLSRALRYDDXNIMTKVHGKRYAYKFDHGI
AQAOLPHPPSSILYKYPSDLPYMGSVHAHQKMFVAPHPALPTVTSFFFAAPNPYW
NSTPGIYPNTRLPSTSHMPSHLGTYI"

ORIGIN

Alignment Scores:

Pred. No.: 3.35e-128 Length: 1291

Score: 1985.50 Matches: 356

Percent Similarity: 96.53% Mismatches: 6

Best Local Similarity: 94.93% Conservative: 12

Query Match: 80.01% Indels: 1

DB: 6 Gaps: 1

US-09-902-772-2 (1-451) x AX657230 (1-1291)

Qy 78 SerArgAsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSer 97

Db 101 TCARGGAACCTCTCTGATGAATGACGATGTGGCCNAGGCGGAGAGATGTGGCGGACGCCA 160

Qy 98 AspAsnValGlyMetAsnTyrGlySerTy-MetGluLysHisIleProProProAsn 117

Db 161 GACACCGTTGGGATGAACCTACGGCAGCTACATGAGGAGAGACATGTCACCCCAAC 220

Qy 118 MetThrThrAsnGluArgArgValIleValProAlaAspProThrIleuTyrSerThrAsp 137

Db 221 ATGACCACGACACAGCGACGAGTATTGTCGCCACGATCTACGCTATGGAGTACAGAC 280

Qy 138 HisValArgGlnTyrLeuGluTyrIleAlaValLysGluTyrGlyLeuProAspValAspIle 157

Db 281 CATGTGGCGAGTGGCTGGAGTGGCGGTGAAGAATATGGCTTCACGACGTCAACATC 340

Qy 158 LeuLeuPheGlnAenIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGln 177

Db 341 TTGTATTTCAGAACATCGATGGGAAGAACTGTGCAAGATGACCAAGAGACGACTTCCAG 400

Tigenix N.V. (BE)
Location/Qualifiers
1. .1372
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
146. .1309
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:29160013"
/db_xref="EMBL:CAD79714.1"
/translation="MVSPPTVGNYSYMEKHPPNNMTNERRVIVPADTLKST
DHVQLEWAWKETGLDPDKNILLFQNDIGKELCKNTKDDFORLTPSNADILLSHLY
LREPLPHLSDDDVKALQNSPRLHARNTGGAAFPNTPSVVPEAQRITTPDLPY
EPPRSWHTGHTPQSKAQPSPFVKPTDQPLDPYQILGPTSSRLANPGSQ
IQWLQFLLELSDSNSCTWEGTNGFNTDPEVAREWGERKSKPNMYDKLARA
LRVYDKNIMTKVGHKYAVKDFPHGIAQALQPHPESSLYKYPDLPYMGSHVHPQ
KMFVAPHPALPYTSSFFPAAPNPYNSPTGGIYPNTRLPTRSHMPSHLQTY"

ORIGIN

Alignment Scores:
Pred. No.: 4,23e-126 Length: 1372
Score: 1926.00 Matches: 355
Percent Similarity: 89.80% Conservative: 6
Best Local Similarity: 88.31% Mismatches: 13
Query Match: 78.81% Indels: 28
DB: 6 Gaps: 2

US-09-902-772-2 (1-451) x AX657234 (1-1372)

QY 78 SerArgAsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSer 97
DB 101 TCARAGAACTCTCTGATGAATGATGAGTGTGCCAAGCGCGGAGATGGTGGCAGGCCA 160
QY 98 AspaenValGlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsn 117
DB 161 GACACCGTGGATGAACACTACGGCAGCTACATGGAGGAGAGACATGCCACCCCAAC 220
QY 118 MetThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAsp 137
DB 221 ATGACCAAGCGGCGAGAGTATATGTCGACAGATCTTACGCTATGGAGTACAGAC 280
QY 138 HisValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIle 157
DB 281 CATGTGCGCAGTGCCTGAGTGGCGGTGAAGAATATGCGCTTCCAGACGTCAACATC 340
QY 158 LeuLeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGln 177
DB 341 TTGTTATTCAGAACATCGATGGGAAGGAATGTGCAAGATGACCAAGGACGACTTCCAG 400
QY 178 ArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlu 197
DB 401 AGGCTCACCCCGAGCTACACGCGGACATCTTCTCTCACATCTCCACTACCTCAGAGAG 460
QY 198 Arg----- 198
DB 461 ACTCCTCTCCACATTTGACTTCAGATGATGTTGATAAAGCCTTACAAAACCTCTCCACGG 520
QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyr 210
DB 521 TTAATGATGTAGAAACACAGGGGGTGCAGCTTTATTTTCCCAATATCTTCAGTATAT 580
QY 211 ProGluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArg 230
DB 581 CCTGAAGCTACGCAAGAAATTAACACTAGGCAGATTACCATATAGAGCCGCCAGGAGA 640
QY 231 SerAlaTrpThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSer 249
DB 641 TCAGCCTGGACCGGTACGCGGCCACCCCGCCAGTCCGAAGCTGCTCAACCATCTCCT 700
QY 250 SerThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGly 269

DB 701 TCCACAGTCCCAAAACTGAAGACCAGCGCTCTCAGTTAGATCCTTATCAGATTCTTGA 760
QY 270 ProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeu 289
DB 761 CCAACAAGTAGCGCTTGCAATCCAGGCACTGGCCAGATCCAGCTTGGCAGTTCCCTC 820
QY 290 LeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGly 309
DB 821 CTGGAGCTCTCTCGGACAGCTCCAACTCCAGCTGCATCACCCTGGGAAGGCAACACGGG 880
QY 310 GluPheLysMetThrAspProAspGluValAlaIaAaGATGTPGlyGluArgLysSerLys 329
DB 881 GAGTTCAGATCAGGATCCCGACGAGGTGGCCGCGCTGGGGAGAGCGGAAGACAAA 940
QY 330 ProAsnMetAsnTyrAspLysLeuSerArgAlaLeuAaGlyTyrTyrAspLysAsnIle 349
DB 941 CCAACATGAATACGATAAGCTACGCGCGCCCTCCGTTACTACTATGACAAGAACATC 1000
QY 350 MetThrLysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGln 369
DB 1001 ATGACCAAGGTCCATGGGAGCGCTACGCTACAGTTCCGACTTCCACGGATCGCCAG 1060
QY 370 AlaLeuGlnProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyr 389
DB 1061 GCCTTCAGCCCAACCCCGCGAGTCACTCTGTACAGTACCCCTCAGACCTCCCGTAC 1120
QY 390 MetSerSerTyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAla 409
DB 1121 ATGGGCTCTATACGCCCCCAGAGAGATGAATTTGTGGCGCCCACTCCCTCCAGGC 1180
QY 410 LeuProValThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThr 429
DB 1181 CTCCCGGTGACATCTTCCAGTTTTTTTGTGCCCCAAACCCATACCTGGAATTCACCACT 1240
QY 430 GlyGlyIleTyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThr 449
DB 1241 GGGGGTATATACCCCAACACTAGGCTCCCAACAGCCATATGCTTCTCATCTGGGCAC 1300
QY 450 TyrTyr 451
DB 1301 TACTAC 1306

Search completed: May 27, 2004, 22:55:53
Job time : 5643 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 27, 2004, 19:25:13 ; Search time 610 Seconds
(without alignments)
3140.881 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTIKALSVDQSLFZ.....IYNTRLPAHMSHLGTYV 451

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Dgapop 6.0 , Delext 7.0

Searched: 337363 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlh
-Q=/Cgn2_1/USPTO_6P001/US09902772/runat 27052004 162840 2136/app query.fasta_1.647
-DB=N Geneseq 29Jan04 -QMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USERS=US09902772 @CGN 1 1 470 @runat 27052004 162840 2136 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=130 -WARN TIMEOUT=30 -THREAS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq 29Jan04.*

1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002s.*
7: geneseqn2003as.*
8: geneseqn2003bs.*
9: geneseqn2003cs.*
10: geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2444	100.0	1447	2 AAX26551	Aax26551 DNA encod
2	2414.5	98.8	1528	8 AAX26552	Aax26552 DNA encod
3	2327.5	95.2	1509	8 ABz68766	Abz68766 Nucleotid
4	2182	89.3	1389	7 ABX08759	Abx08759 Angiogene
5	2182	89.3	3166	7 ABX76295	Abx76295 Lung canc
6	2182	89.3	3166	9 ADB75274	ADB75274 Prostata
7	1955.5	80.0	1291	8 ABz68767	Abz68767 Nucleotid
8	1926	78.8	1372	8 ABz68769	Abz68769 Nucleotid

9	1807.5	74.0	1219	8	ABz68768	Abz68768 Nucleotid
10	1624.5	66.5	1359	8	ADA02692	Human FLI
11	1624.5	66.5	1359	9	ADB72430	Human FLI
12	1624.5	66.5	2938	6	AAQ50644	AAQ50644 Human ca
13	1624.5	66.5	2957	2	ABV94304	ABV94304 Breast ca
14	1624.5	66.5	2957	6	ABK84139	Human CDN
15	1624.5	66.5	2957	8	ADA02691	Human FLI
16	1624.5	66.5	2957	9	ADB72429	Human FLI
17	1613.5	66.0	1359	8	ADA02689	Mouse FLI
18	1613.5	66.0	1359	9	ADB72427	Mouse FLI
19	1613.5	66.0	1729	8	ADA02688	Mouse FLI
20	1613.5	66.0	1729	9	ADB72426	Mouse FLI
21	1599	65.4	2954	4	AAH02915	Human she
22	1506.5	61.6	1890	2	AAQ50662	Human Hum
23	965	39.5	1431	6	ABs73253	ABs73253 DNA encod
24	948	38.8	567	4	AAI20932	Probe #10
25	948	38.8	567	4	ABA66003	Human foe
26	948	38.8	567	4	AAI46176	Probe #14
27	948	38.8	567	4	ABA48124	Human bre
28	948	38.8	567	4	ABA33090	Probe #11
29	948	38.8	567	4	AAK40160	Human bon
30	948	38.8	567	4	AAK14426	Human bra
31	948	38.8	567	4	ABS39742	Human liv
32	948	38.8	567	5	AAI06643	Probe #66
33	948	38.8	567	6	ABS14215	Human gen
34	863	35.3	473	4	ABA68128	Human foe
35	863	35.3	473	4	AAI48339	Probe #17
36	863	35.3	473	4	ABA50193	Human bre
37	863	35.3	473	4	ABA35152	Probe #13
38	863	35.3	473	4	AAK42265	Human bon
39	863	35.3	473	4	AAK16512	Human bra
40	863	35.3	473	4	ABS41873	Human liv
41	863	35.3	473	5	AAI08696	Probe #86
42	833.5	34.1	73725	9	ADA02690	Human FLI
43	833.5	34.1	73725	9	ADB72428	Human FLI
44	813.5	33.3	72732	8	ADA02687	Mouse FLI
45	813.5	33.3	72732	9	ADB72425	Mouse FLI

ALIGNMENTS

RESULT 1

AAX26551

ID AAX26551 standard; DNA; 1447 BP.

XX AC AAX26551;

XX AC AAX26551;

DT 14-JUN-1999 (first entry)

XX DT 14-JUN-1999 (first entry)

XX DE DNA encoding chicken C-11 protein.

XX DE DNA encoding chicken C-11 protein.

XX KW Chicken; C-11 protein; cell calcification inhibiting activity;

XX KW cell calcification inhibiting agent; c-erg protein; arthritis deformans;

XX KW ossification; spinal column ligament; ss.

XX OS Gallus sp.

XX OS Gallus sp.

XX FH Key

XX FH Key

FT CDS

FT CDS

XX FT 63..1418

XX FT /*tag= a

XX PN JP11075871-A.

XX PN JP11075871-A.

PD 23-MAR-1999.

XX PD 23-MAR-1999.

XX PF 29-MAY-1998; 98JP-00166076.

XX PF 29-MAY-1998; 98JP-00166076.

XX PR 18-JUN-1997; 97US-00878177.

XX PR 18-JUN-1997; 97US-00878177.

XX PR 20-JUN-1997; 97US-0050297P.

XX PR 20-JUN-1997; 97US-0050297P.

XX PA (CHUS) CHUGAI PHARM CO LTD.

XX PA (CHUS) CHUGAI PHARM CO LTD.

XX PA (UTYPE-) UNIV PENNSYLVANIA.

XX PA (UTYPE-) UNIV PENNSYLVANIA.

DR WPI; 1998-257708/22.
 DR P-PSDB; AAY01520.
 XX An active protein for inhibiting cell calcification - useful for
 PT measuring the calcification of a cell, for diagnosing arthritis deformans
 PT or ossification of spinal column ligament.
 PS Disclosure; Page 7-8; 15pp; Japanese.
 XX The present sequence encodes a chicken C-11 protein which has cell
 CC calcification inhibiting activity. The specification also describes a
 CC cell calcification inhibiting agent containing c-erg protein (AAY01521).
 CC The proteins are used for measuring the calcification of a cell, for
 CC diagnosing arthritis deformans or ossification of spinal column ligament
 XX Sequence 1447 BP; 440 A; 374 C; 317 G; 316 T; 0 U; 0 Other;
 SQ

Alignment Scores: 2.01e-199 Length: 1447
 Pred. No.: 2444.00 Matches: 451
 Score: 2444.00 Conserv: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 100.00% Gaps: 0
 DB: 2

US-09-902-772-2 (1-451) x AAX26551 (1-1447)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
 Db 63 ATGCAGAGCACTATTAGGAAGCAATTATCAGTGTGTGAGTGAACCGATCTTGTGTGAG 122
 Qy 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
 Db 123 TGTGCTACGGATCGCCACCTTGTCAAGACAGAAATGACAGCTCTCTCCAGTGAA 182
 Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
 Db 183 TATGGCAAAATCAATCAAGATGAGCCGCGCTTCCCGAGGAGACTGGTTATCACAGCCC 242
 Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 Db 243 CGGCCAGAGTTACATTAAGATGGAGTGTAAACCAACCCAGGTTAATGGTCAAGGAAT 302
 Qy 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
 Db 303 TCACCTGATGACTCGCGTGGCAAAAAGGAGGAAATGGTTAGCAGTTCAGACAAATGTT 362
 Qy 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120
 Db 363 GGGATGAACCTATGGAGCTACATGGAGAGAGAGCATATTCCGCTCCCAATATGACAAAC 422
 Qy 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
 Db 423 AATGAACGAAGAGTTATTGTGCCAGCAGATCCTACGTTATGGAGCACAGACCATGTACGG 482
 Qy 141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
 Db 483 CAGTGGCTGGAGTGGGAGTGAAGAGTATGGTCTTCCAGACGTGGACATCTTGTCTTC 542
 Qy 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
 Db 543 CAGACATGATGGGAAGAGTTGTGTAAATGACCAAAAGATGACTTCCAGAGACTCAAG 602
 Qy 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyVala 200
 Db 603 CCGAGCTATACGCGAGATATCTCTGTCAACCTACACTACCTACCTACAGAGAGAGAGGCC 662
 Qy 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220
 Db 663 ACTTTTATTTTCCAAATACATCACTTTACCCAGAGCAACGCAAGGAATTAACAACAGG 722
 Qy 221 ProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThr 240

Db 723 CCAGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGAGCTCAGCCATCCCACT 782
 Qy 241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260
 Db 783 CAGTCAAAAGCTACCCCAACCATCATCTTCAACAGTGGCCAAACAGAAAGACCCAGCTCT 842
 Qy 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
 Db 843 CAGTTAGATCTTATCAGATTTCTTGGACCGACAGCAGCGCTCTTGCAAATCCAGGGAGT 902
 Qy 281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeuSerSerSerSerSerSerAsn 300
 Db 903 GGGCAGATACAGCTATGGCAGTTCTCTAGTGGAGTCTTCTCGACAGCTCCAACTCCCAAC 962
 Qy 301 CysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320
 Db 963 TGCATCACCTGGAGGGGACAAATGGGGAGTTCAAGATGACAGACCCCTGTGAAGTGGCT 1022
 Qy 321 ArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
 Db 1023 CGGCGTTTGGGAGAGAGAGAAACCTTAACATGAACATATGACAACTCAGCCGTGCA 1082
 Qy 341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
 Db 1083 CTTCGCTACTACTATGACAAAATATATGACTAAAGTTTCATGTTAAACGCTATGCTTAC 1142
 Qy 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMet 380
 Db 1143 AAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCACCCTCCAGAAATCATCCATG 1202
 Qy 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
 Db 1203 TACAATACCATCCATCAGACTCCCTCATGATGATTCCTACCATGCACACCCCCAGAGATG 1262
 Qy 401 AsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhePheAlaAla 420
 Db 1263 AACTTTGTAGTCCCATCCCTCTGCTTTGGCCGTAACCTCATCCAGCTTTTGTCTGCC 1322
 Qy 421 ProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAla 440
 Db 1323 CCTAATCCATACCTGGAATTCACCAACTGGAGCATCTACCCCAATACCCAGGCTGCCAGCT 1382
 Qy 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1383 GCTCATATGCTTCCCATCTTGGCACCTACTACTAC 1415

RESULT 2
 AAX26552
 ID AAX26552 standard; DNA; 1528 BP.
 XX
 AC AAX26552;
 XX
 DT 14-JUN-1999 (first entry)
 XX
 DE DNA encoding chicken c-erg protein.
 KW Chicken; C-11 protein; cell calcification inhibiting activity;
 KW cell calcification inhibiting agent; c-erg protein; arthritis deformans;
 KW ossification; spinal column ligament; ss.
 XX
 OS Gallus sp.
 XX
 FH Key Location/Qualifiers
 CDS 63..1499
 FT /*tag = a
 XX
 PN JP11075871-A.
 XX
 PD 23-MAR-1999.
 XX
 PF 29-MAY-1998; 98JP-00166076.
 XX
 PR 18-JUN-1997; 97US-00878177.

PR 20-JUN-1997; 97US-0050297P.
 XX (CHUS) CHUGAI PHARM CO LTD.
 PA (UYPE-) UNIV PENNSYLVANIA.
 XX WPI; 1999-257708/22.
 DR P-PSDB; AAY01521.
 XX An active protein for inhibiting cell calcification - useful for
 PT measuring the calcification of a cell, for diagnosing arthritis deformans
 PT or ossification of spinal column ligament.
 XX
 XX Disclosure; Page 8-9; 15pp; Japanese.
 XX
 CC The present sequence encodes a chicken c-erg protein. The specification
 CC also describes a chicken C-11 protein (AAY01520) which has cell
 CC calcification inhibiting activity and a cell calcification inhibiting
 CC agent containing c-erg protein. The proteins are used for measuring the
 CC calcification of a cell, for diagnosing arthritis deformans or
 CC ossification of spinal column ligament
 XX
 XX Sequence 1528 BP; 464 A; 395 C; 330 G; 339 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 7, 26e-197 Length: 1528
 Score: 2414.50 Matches: 450
 Percent Similarity: 94.14% Conservative: 0
 Best Local Similarity: 94.14% Mismatches: 1
 Query Match: 98.79% Indels: 27
 DB: 2 Gaps: 1

US-09-902-772-2 (1-451) x AAX26552 (1-1528)

QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerIleuPheGlu 20
 Db 63 ATGGCAAGCACTATTATGGAAGAGCATTTATCAGTGGTGCAGTGAAGACCACTCTTTTTCAG 122
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGlu 40
 Db 123 TGTGCTTCAGGATCCGCCACCTTCGCAAGAGCAGAAATGACAGCTCTCTTCCAGTGAA 182
 QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
 Db 183 TATGGCAAAACATCAAGATGAGCCGCGGTCCCGAGCAGGACTGGTTATCAAGCC 242
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 Db 243 CCGGCCAGAGTTACCATTAAGATGGAGTGTAAACCCAAACCCAGTTATGGTCAAGGAT 302
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
 Db 303 TCACCTGATGACTGCAGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTCAGACAATGTT 362
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
 Db 363 GGGATGAACATATGGAAGCTACATGGAAGAGAGCATATTCGCCCTCCAAATATGACAACC 422
 QY 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
 Db 423 AATGAACGAGAGTATTATGTCAGCAGATCTCTAGCTTATGGAGCAGACCATGTACGG 482
 QY 141 GlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
 Db 483 CAGTGGCTGGAGTGGGAGTGAAGGAGTATGCTCTTCAGAGCTGGACATCTTGTGTTC 542
 QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
 Db 543 CAGAACATATGATGGGAAGAGTGTGTAAATATGACCAAAAGATGATTCAGAGATCTCAG 602
 QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg----- 198
 Db 603 CCGAGCTATAACGACATATCTCTCTGTTCACACCTACACTACCTCAGAGAGACTCTCTT 662

QY 198 ----- 198
 Db 663 CCACATTTGACTTCAGATGATGTTGATAAGGCTTTACAAAACCTCTCCAGGTTAATGAT 722
 QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213
 Db 723 GCTAGAAACACAGGAGGAGCCACATTTATTTTCCAAATACATCAGTTTATCCAGAGCA 782
 QY 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrp 233
 Db 783 ACGCAAGAATTAACAAACAGCCAGATTACCTTATGAGCAAGCAGGAGATCACCGTGG 842
 QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 253
 Db 843 ACGAGTCACAGCCATCCCACTCAGTCAAAAGTACCCCAACCATCATCTTCAACAGTGCC 902
 QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273
 Db 903 AAAACAGAGACAGCGTCTCAGTTAGATCCTTATCAGATTCTTGGACCGACAGACAGC 962
 QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeu 293
 Db 963 CGTCTTTGCAATCCAGGGAGTGGGAGATACAGCTATGCGAGTTCTTACTGGAGCTTCTG 1022
 QY 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
 Db 1023 TCGGACAGCTCCCACTCCCACTGTCATCACTGGGAGGCGACAAATGGGGAGTTCAAGATG 1082
 QY 314 ThrAspProAspGluValAlaArgTyrGlyGluArgLysSerLysProAsnMetAsn 333
 Db 1083 ACAGACCTTGATGAGTGGCTGGGAGTGGGAGAGGAAAGCAAACTTACATGAAC 1142
 QY 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysVal 353
 Db 1143 TATGACAAACTCAGCGTGCACCTCGCTACTACTATGACAAAATATTATGACTAAAGTT 1202
 QY 354 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
 Db 1203 CATGGTAAACGCTATGCCCTACAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGCCT 1262
 QY 374 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393
 Db 1263 CACCTCCAGATCATCCATGATGACAAATACCCATCAGACCTCCCTTACATGAGTCTCTAC 1322
 QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413
 Db 1323 CATGCAACCCCAAGATGAACCTTGTAGCTCCCATCCCTCTGCTTTGCCCGTAACC 1382
 QY 414 SerSerSerPhePheAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyr 433
 Db 1383 TCATCCAGCTTTTGTGCTGCCCTTATCCATCTGGAATTCACCAACTGGAGCATCTAC 1442
 QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1443 CCCAATACAGGCTGCCAGCTGCTCATATGCTTCCCATCTTCCCATCTTGGCACCTACTAC 1496

RESULT 3
 ABZ68766 standard; DNA; 1509 BP.
 ID ABZ68766
 XX AC ABZ68766;
 XX
 XX 16-MAY-2003 (first entry)
 DT
 XX
 DE Nucleotide sequence of human ERG splice variant C-1-1.
 XX
 KW Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
 KW C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
 KW cartilage disease; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 PH

```

CDS      67..1446
          /*tag= a
          /product= "ERG splice variant C-1-1"
WO2003000724-A2.
03-JAN-2003.
08-MAR-2002; 2002MO-EP002605.
25-JUN-2001; 2001US-0300756P.
(TIGE-) TIGENIX NV.
Luyten F, De Bari C, Dell'accio F;
WPI; 2003-221398/21.
P-PSDB; ABP97695.
XX Novel splice variants of the Ets-related gene (ERG) transcription factor
PT , useful for preventing or treating bone-related or cartilage-related
PT diseases or Ets-related gene associated tumor.
XX Claim 9; Fig 4; 60pp; English.
PS The present sequence encodes splice variant C-1-1 of human ERG. ERG
CC belongs to the Ets family of transcription factors. The specification
CC describes C-1-1 (characterised by the absence of the 81 bp fragment, and
CC the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA
CC sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp
CC fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA
CC sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and
CC 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by
CC the absence of the 218 bp fragment, and the presence of the 72 bp and 81
CC bp fragments of the ERG-3 cDNA sequence). The splice variant
CC polynucleotides are useful for stabilizing the phenotype of chondrocytes
CC or cells with a chondrocytes-like phenotype, for preventing de-
CC differentiation of articular chondrocytes during in vitro culture
CC expansion, for preventing or treating bone-related or cartilage-related
CC diseases or Ets-related gene associated tumour, or as marker of
CC progenitor cells of stable chondrocytes. Cells expressing the
CC polynucleotide are useful for producing or repairing cartilage tissue in
CC a mammal
XX
SQ Sequence 1509 BP; 389 A; 472 C; 359 G; 289 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      1.99e-189      Length:      1509
Score:          2327.50      Matches:      427
Percent Similarity: 96.90%      Conservative: 10
Best Local Similarity: 94.68%      Mismatches: 13
Query Match:    95.23%      Indels:      1
DB:             8      Gaps:      1

US-09-902-772-2 (1-451) x ABZ68766 (1-1509)
QY      2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
Db      91 GCAGCTCATATCAAGGAGCCCTTATCAGTTGTGAGTGGAGCCAGTCTGTTTGTAGTGT 150
QY      22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyr 41
Db      151 GCCTACGGACGCCACACCTGGCTAAGACAGATGACCGGCTCTCTCCAGCGACTAT 210
QY      42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProPro 61
Db      211 GGACAGACTTCCAGATGAGCCACGCGCTCCCTCAGCAGGATGGCTGTCTCAACCCCA 270
QY      62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db      271 GCCAGGGTCAACATCAAAATGGATGTAACCTAGCCAGGATGAATGGCTCAAGGAAGTCT 330
QY      82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGly 101

```

```

Db      331 CCTGATCAATGCAGTGTGGCCAAAGCGGGAGAGATGTTGGCGCACACCGTTGGG 390
QY      102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProProProAsnMetThrThrAsn 121
Db      391 ATGAATACGCGCAGCTACATGGAGGAGAGACATGCCACCCCAACACATGACCCAGAAC 450
QY      122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
Db      451 GAGCGCAGAGTTATCGTCCAGCAGATCTTACGCTATGGAGTACAGACCATGTGGCGCAG 510
QY      142 TrpLeuGlnTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
Db      511 TGGCTGGAGTGGCGGTGAAGATATGGCTTCCAGACGCTCAACATCTTGTATTCCAG 570
QY      162 AsnIleAspGlyLysGluLeuLysMetThrLysAspAspPheGlnArgLeuThrPro 181
Db      571 AACATCGATGGAGAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCC 630
QY      182 SerTyrAsnAlaAspIleLeuLeuSerHisHisIleValLeuArgGluArgGlyAlaThr 201
Db      631 AGCTACAACGCGCAGATCTCTCTCATCTCCACTACCTCAGAGAGAGGGGTGCAGCT 690
QY      202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
Db      691 TTTATTTTCCAAATATCTTCACTATATCTTGAAGCTACGCAAGAAATATCACTAGGCCA 750
QY      222 AspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThr--- 240
Db      751 GATTATCATATGAGCCCCCAGGAGATCAGCTTGGACCGGTCCAGCCACCCACGCC 810
QY      241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260
Db      811 CAGTCGAAAGTGTCTCAACCATCTCTTCCAGTGGCCCAAACTGAAGACAGCGTCTCT 870
QY      261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
Db      871 CAGTTAGATCTTATCATGATCTTGGACCAAGTAGTACCGCTTGCATATCCAGGCA 930
QY      281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsn 300
Db      931 GCGCGCTGGGAGAGCGGAGAACCAACCACTAGAACTACGATAGCTCAGCCGCCGCC 1110
QY      341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
Db      1111 CTCCTTTACTACTATGACAGAACATCATGACCAAGTCTTCCATGGAGAGCGCTACGCTAC 1170
QY      361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerSerMet 380
Db      1171 AAGTTTCGACTTCCAGGGATCGCCAGCGCCCTCCAGCCCAACCCCGAGTCATCTCTG 1230
QY      381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
Db      1231 TACAGTACCTCTCAGACTCCCTGATCATGGCTCTCTATCAGCCCAACCCAGAGATG 1290
QY      401 AsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaAla 420
Db      1291 AACTTGTGGCGCCCAACCTCCAGCCCTCCCGGTGACATCTTCCAGTCTTTTCTGCTGC 1350
QY      421 ProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAla 440
Db      1351 CCAACCCCAATCTGGAATTCACCAACTGGGGGTATATACCCCAACACTAGGCTCCCCACC 1410
QY      441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451

```

Db 1411 AGCCATATGCTTCTCATCTGGGCACTTACTAC 1443

RESULT 4

ABX08759

ID ABX08759 standard; cDNA; 1389 BP.

XX ABX08759;

XX 21-JAN-2003 (first entry)

XX DT

XX Angiogenesis-associated human polynucleotide sequence #21.

DE XX Human; angiogenesis-associated transcript; angiogenesis;

XX KW angiogenesis-associated disease; cancer; cytostatic; gene therapy; gene;

XX KW ss.

XX OS Homo sapiens.

XX PN WO200279492-A2.

XX PD 10-OCT-2002.

XX PF 14-FEB-2002; 2002WO-US004915.

XX PR 14-FEB-2001; 2001US-00784356.

XX PR 22-FEB-2001; 2001US-00791390.

XX PR 19-APR-2001; 2001US-0285475P.

XX PR 03-AUG-2001; 2001US-0310025P.

XX PR 13-NOV-2001; 2001US-0350666P.

XX PR 29-NOV-2001; 2001US-0334244P.

XX PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX PI Murray R, Glynn R, Watson SR, Aziz N;

XX DR WPI; 2003-040681/03.

XX DR P-PSDB; ABU03476.

XX PT Detecting angiogenesis-associated transcript in a cell for diagnosing and
 PT treating cancer by contacting a sample with a polynucleotide that
 PT exhibits changes in expression level as a function of time in tissue
 PT undergoing angiogenesis.

PS Example 2; Page 200; 291pp; English.

XX CC The present invention relates to methods and compositions for detecting
 CC an angiogenesis-associated transcript in a cell in a patient. The method
 CC involves contacting a biological sample from the patient with a
 CC polynucleotide that selectively hybridises to a sequence at least 80%
 CC identical to any of the angiogenesis-associated human polynucleotide
 CC sequences given in the specification. These angiogenesis-associated
 CC polynucleotide sequences comprise genes that exhibit changes in
 CC expression levels as a function of time in tissue undergoing
 CC angiogenesis. The method and the polynucleotide sequences of the
 CC invention are useful for diagnosing and treating angiogenesis and
 CC angiogenesis-associated diseases e.g. cancer. The polynucleotide
 CC sequences are also useful in the gene therapy of such disorders. The
 CC angiogenesis-associated proteins encoded by the polynucleotide sequences
 CC are useful as a vaccine for therapeutic and prophylactic immunisation.
 CC ABX08759-ABX08853 represent angiogenesis-associated polynucleotide
 CC sequences

SQ Sequence 1389 BP; 359 A; 440 C; 323 G; 267 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,048-177	Length:	1389
Score:	2182.00	Matches:	407
Percent Similarity:	91.72%	Conservative:	14
Best Local Similarity:	88.67%	Mismatches:	24
Query Match:	89.28%	Indels:	14
DB:	7	Gaps:	4

US-09-902-772-2 (1-451) x ABX08759 (1-1389)

QY	2	AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys	21
Db	25	GCAGCTCATATCAAGAAAGCCCTTATCAGTTGTAGTGAGGACCACTGCTTTGTGAGGT	84
QY	22	AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr	41
Db	85	GCCTACGGGAACGCCACACCTGGCTAAGACAGAGATGACCGGTCTCTCCACGCACTAT	144
QY	42	GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProPro	61
Db	145	GGACAGACTTCCAGATGAGCCCGCTCCCTCAGCAGGATTGGCTGTCTCAACCCCA	204
QY	62	AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer	81
Db	205	GCCAGGGTCACCATCAAAATGAAATGAATTAACCTAGCCAGGTGAATGGCTCAAGCACTCT	264
QY	82	ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerSerSerValGly	101
Db	265	CTGTGATGAATGCGAGTGTGCCCAAGCGGGAAGATGGTGGCAGCCAGACCGTTGGG	324
QY	102	MetAsnTyrGlySerTyrMetGluLysHisIleProProProAsnMetThrThrAsn	121
Db	325	ATGAACCTACGGCAGCTACATGGAGGAGACACATGCCACCCCAACATGACCACCAAC	384
QY	122	GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln	141
Db	385	GAGCGCAGAGTTATCGTCCGACGAGATCTTACGCTTAGAGTACAGACCATGTGGCGAG	444
QY	142	TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln	161
Db	445	TGGCTGGAGTGGCGGTGAAGAATATGGCTTCCAGACGTCAACATCTTTATTCCAG	504
QY	162	AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro	181
Db	505	AACATCGATGGAGGAACTGTGCAAGATGCCAAGGACGACTTCCAGAGCTCACCCTCC	564
QY	182	SerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr	201
Db	565	AGCTACAAACGGCAGCATCTCTCTCATCTCCACTACCTACAGAGACTCTCTTCCA	624
QY	202	PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro	221
Db	625	CATTGG-----ACITCA-----GATGATTTGATAAGCCTTACAAAACCTTCCA	669
QY	222	-----AspLeuProTyrGluGlnAlaArgArgSerAlaTrp	233
Db	670	CGGTTATGTCATGCTAGAAACACAGATTTACCATATGAGCCCCCAGAGATCAGCCTGG	729
QY	234	ThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrVal	252
Db	730	ACCGGTCAGGCCACCCACGCCCGCAGTCAAGAGCTCTCAACATCTCTTCCACAGTG	789
QY	253	ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer	272
Db	790	CCCAAACTGAAGACCCAGCGTCTCTAGTATGCTTATCAGATTCTTGGACCAACAAGT	849
QY	273	SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu	292
Db	850	AGCCGCTTGCATAATCCAGGAGTGGCCAGATCAGCTTTGGCAGATTCTCTCTGGAGCTC	909
QY	293	LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys	312
Db	910	CTGTCCGACAGCTCCAACCTCCAGCTGCATCACCTGGGAAGGCCACCAACGGGGATTCAAG	969
QY	313	MetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMet	332
Db	970	ATGACGGATCCGACAGAGTGGCGCGCTGGGAGAGCGGAGAGCAACCAACATG	1029
QY	333	AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLys	352
Db	1030	AACTACGATAGCTCAGCGCGCCCTCCGTTACTTACTTATGACAAGAACATCATGACCAAG	1089

Qy 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
 Db 1090 GTCCATGGGAAGCTACGCTACAGTTCGACTTCACGGGATGCGCCAGGCCCTCCAG 1149
 Qy 373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
 Db 1150 CCCACCCCGGAGTCACTCTGTACAGTACCCCTCAGACTCCCGTACATGGGCTCC 1209
 Qy 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
 Db 1210 TATCACGCCACCCACAGAGATGAATTTGTGGCGCCGCCACCTCCAGGCCCTCCCGGTG 1269
 Qy 413 ThrSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIle 432
 Db 1270 ACATCTTCCAGTCTTTTGTGCGCCCAACCCACTGGAATTCACCACTGGGGGTATA 1329
 Qy 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1330 TACCCACACATAGCTCCCGCCACGACCATATGCTTCTCATCTGGGCACTTACTAC 1386
 RESULT 5
 ABX76295
 ID ABX76295 standard; DNA; 3166 BP.
 XX
 AC ABX76295;
 DT 02-APR-2003 (first entry)
 XX
 DE Lung cancer-associated polynucleotide #159.
 XX
 KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX
 OS Unidentified.
 XX
 PN WO200286443-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 18-APR-2002; 2002WO-US012476.
 XX
 PR 18-APR-2001; 2001US-0284770P.
 PR 10-MAY-2001; 2001US-0290492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350665P.
 PR 29-NOV-2001; 2001US-0334370P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Aziz N, Murray R;
 XX
 XX WPI; 2003-093161/08.
 DR P-PSDB; ABU56566.
 XX
 PT Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer.
 XX
 XX Claim 22; Page 310-311; 453pp; English.
 PS
 CC The invention relates to a method for detecting a lung cancer-associated
 CC transcript in a cell from a patient, comprising contacting a biological
 CC sample from the patient with a polynucleotide that selectively hybridises
 CC to a sequence that is at least 80 % identical to a gene that exhibits
 CC increased or decreased expression in lung cancer samples. Lung cancer-
 CC associated polynucleotides and polypeptides are used for identifying a
 CC compound that modulates a lung cancer-associated polypeptide, for
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung

CC cancer in a patient and for treating a mammal having lung cancer by
 CC administering a modulatory compound identified. The methods are useful
 CC for treating lung cancer, such as small cell lung cancer, non-small cell
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
 CC for diagnostic purposes and as targets for screening for therapeutic
 CC compounds that modulate lung cancer, such as antibodies. Sequences
 CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
 CC invention
 XX
 SQ Sequence 3166 BP; 915 A; 770 C; 768 G; 713 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.55e-176 Length: 3166
 Score: 2182.00 Matches: 407
 Percent Similarity: 91.72% Conservative: 14
 Best Local Similarity: 88.67% Mismatches: 24
 Query Match: 89.28% Indels: 14
 Ds: 7 Gaps: 4
 US-09-902-772-2 (1-451) x ABX76295 (1-3166)
 Qy 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
 Db 281 GCAGCTCATATCAGGAGAGCCCTTATCAGTTGTGAGTGAGGACCGTGTGTGTGAGTGT 340
 Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41
 Db 341 GCCTACGGAGCCACACCTGGCTAAGACAGAGATGACCGGTCTCTCTCCAGGACATAT 400
 Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61
 Db 401 GGACAGACTTCCAGATGAGCCACGCGTCCCTCAGCAGGATGGTGTCTCAACCCCA 460
 Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
 Db 461 GCCAGGCTCACCATCAAAATGGAATGTAACCTAGCCAGGTGAATGGCTCAAGGAACCT 520
 Qy 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly 101
 Db 521 CCTGATGAATGCAGTGTGGCCAAAGCGGGAAGATGGTGGGAGCCGACACACCGTTGG 580
 Qy 102 MetAsnTyrGlySerTyrMetGluGlnLysHisIleProProProAsnMetThrThrAsn 121
 Db 581 ATGAATACGGCAGCTACATGAGGAGAGAGACATGCCACCCCAACATGACCCAGAAC 640
 Qy 122 GluArgValValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141
 Db 641 GAGCGCAGAGTTATCGTGCACAGATCTCTACGCTATGGAGTACAGACCATGTGGGCGAG 700
 Qy 142 TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
 Db 701 TGGCTGGAGTGGGCGGTGAAGATATGGCTTCCAGACGTCACACATCTGTATTTCAG 760
 Qy 162 AsnIleAspGlyLysGluLeuLysCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
 Db 761 AACATCGATGGGAGGAACTGTGAGATGACAGGAGACGACTTCAGAGGCTCACCC 820
 Qy 182 SerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
 Db 821 AGCTACACGCCGACATCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 880
 Qy 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
 Db 881 CATTTG-----ACTTCA-----GATGATGTTGATAAGCCCTTACAAACTCTCCA 925
 Qy 222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTyr 233
 Db 926 CGGTTAATGCTAGTAACACAGATTTACCATATGAGCCCCCAGGAGATCAGCGCTGG 985
 Qy 234 ThrSerHisSerHisProThr-----GlnSerLysAlaThrGlnProSerSerThrVal 252

Db 986 ACAGTACAGGCGACCCAGCCAGCCAGTGAAGCTGCTCAACATCTCTCCACAGTG 1045
 Qy 253 ProlysthrGluAspGlnArgProGlnLeuAspProTyrGlnLeuGlyProThrSer 272
 Db 1046 CCCAAACTGAAGACAGGCTCTCAGTATAGATCTTATCAGATTTCTGGACCAAGT 1105
 Qy 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuGluLeu 292
 Db 1106 AGCGGCTTCAATCCAGGAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCGGAGCTC 1165
 Qy 293 LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys 312
 Db 1166 CTGTGGACAGCTCACTCAGCTGATCATCCTGGGAGGCCACCAACGGGAGTTCAAG 1225
 Qy 313 MetThrAspProAspGluAlaAlaArgArgTyrGlyGluArgLysSerLysProAsnMet 332
 Db 1226 ATGACGATCCGACAGGAGTGGCCGCTGGGAGAGCGGAGAGCAACCAACATG 1285
 Qy 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 352
 Db 1286 AACTACATAGCTCAGCGGCGCTCCGTACTACTATGACAGACATCATGACCAAG 1345
 Qy 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
 Db 1346 GTCCATGGGAGCGCTACGCTACAGTTCGATTCACGGGATCGCCAGGCGCTCCAG 1405
 Qy 373 ProHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
 Db 1406 CCCACCCCGGAGTCTCTCTGTAACAGTACCCCTCAGACCTCCCGTACATGGGCTCC 1465
 Qy 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
 Db 1466 TATCACGCCCCACCCACAGAGAGTGAATCTTTGGCGGCCCCACCTCCAGCGCTCCCGGTG 1525
 Qy 413 ThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIle 432
 Db 1526 ACAFTTCCAGTATTTTGTGTCGCCCAACCCATACCTGGAATTCACCACTGGGGGTATA 1585
 Qy 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1586 TACCCCAACACTAGGCTCCCAACAGCCATATGCTTCTCATCTGGGCACTTACTAC 1642
 RESULT 6
 ADB75274
 ID ADB75274 standard; cDNA; 3166 BP.
 AC ADB75274;
 DT 04-DEC-2003 (first entry)
 DE Prostate cancer marker cDNA.
 XX Prostate; cancer; cytostatic; gene therapy; marker; ss.
 OS Homo sapiens.
 PN WO2003009814-A2.
 XX 06-FEB-2003.
 PF 25-JUL-2002; 2002WO-US023913.
 XX 25-JUL-2001; 2001US-0307982P.
 PR 22-AUG-2001; 2001US-0314356P.
 PR 25-SEP-2001; 2001US-0325020P.
 PR 12-DEC-2001; 2001US-0341746P.
 PR 05-MAR-2002; 2002US-0362159P.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;
 PI Hoersher S, Kamatkar S, Woney AM, Glatt K, Zhao X, Anderson D;

XX WPI; 2003-248033/24.
 XX New nucleic acid molecule, useful for diagnosing or treating prostate cancer.
 PT Disclosure; SEQ ID NO 98; 99pp; English.
 PS The invention relates to newly discovered cancer markers associated with the cancerous state of prostate cells. Also disclosed is a method of assessing whether a patient is afflicted with prostate cancer. The method of the invention involves assessing whether a patient is afflicted with prostate cancer by comparing the level of expression of a marker in a patient sample and the normal level of expression of the marker in a control non-prostate cancer sample, where a significant increase in the level of expression of the marker in the patient sample and the normal level indicates that the patient is afflicted with prostate cancer.
 CC Nucleic acids of the invention are useful for diagnosing or treating prostate cancer, and may be useful in gene therapy. Sequences given in ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 3166 BP; 915 A; 770 C; 768 G; 713 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 1.55e-176 Length: 3166
 Score: 2182.00 Matches: 407
 Percent Similarity: 91.72% Conservative: 14
 Best Local Similarity: 88.67% Mismatches: 24
 Query Match: 99.28% Indels: 14
 DB: Gaps: 4
 US-09-902-772-2 (1-451) x ADB75274 (1-3166)
 Qy 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
 Db 281 CGAGCTCATATCAGAAAGCCCTTATCAGTTGTGAGTGAGGACCACTGTTGTTGAGTGT 340
 Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41
 Db 341 GCCTACGGAAACGCCACACTGGCTAAGACAGAGATGACGGCTCTCTCCACGCACTAT 400
 Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAsnTrpLeuSerGlnProPro 61
 Db 401 GGACAGACTTCCAGATGAGCCACGCGTCCCTCAGCAGGATGCTGTCTCACACCCCA 460
 Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
 Db 461 GCCAGGGTCACCATCAAAATGGAATGTAACCTAGCCAGGTGAATGGCTCAAGGAATCT 520
 Qy 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly 101
 Db 521 CCTGATGAATGAGTGTGGCCAAAGCGGGAAGATGGTGGGCGAGCCACACCGTTGGG 580
 Qy 102 MetAsnTyrGlySerTyrMetGluGlyLysHisIleProProProAsnMetThrThrAsn 121
 Db 581 ATGAATACCGCAGCTACATGAGGAGAGACACATGCCACCCCAACATGACACCAAG 640
 Qy 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
 Db 641 GAGGCGCAGAGTTATCGTGCACAGCATCTACGCTATGAGGTACAGACCATGTGGCGCAG 700
 Qy 142 TrpLeuGluTrpAlaValLysGlyLysGlyLeuProAspValAspIleLeuLeuPheGln 161
 Db 701 TGGCTGGAGTGGCGGTGAAGAATATGGCTTTCAGACGCTCAACATCTGTATTATCCAG 760
 Qy 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
 Db 761 ARACTGATGGGAGGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCCCC 820
 Qy 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201

Db 341 TTGTTATTCAGAACATCGATGGGAAGAACTGTGCAAGATGACCAAGGAGCACTTCCAG 400
 QY 178 ArgLeuThrProSerTyAsnAlaSpileuLeuSerHisleuHisTyLeuArgGlu 197
 Db 401 AGGCTCACCCAGCAGTCAAGCCGACATCTTCTTCACATCTCCACTACCTCCAGAG 460
 QY 198 ArgGlyAlaThrPheilePheProAsnThrSerValTyProGluAlaThrGlnArgile 217
 Db 461 AGGGGTGAGCTTTTATTTTCCAAATACTTCCAGTATATCTGAAGCTACGCAAGAAAT 520
 QY 218 ThrThrArgProAspLeuProTyGluGlnAlaArgSerAlaThrThrSerHisSer 237
 Db 521 ACACTAGCCAGATTTTACCATATGAGCCGCCAGGAGATCAGCTCGACCGGTACCGC 580
 QY 238 HisProThr---GlnSerLeuAlaThrGlnProSerSerThrValProLysThrGlu 256
 Db 581 CACCCACGCCCGCAGTCAAGAGTCTCAACATCTCTTCCACAGTGCACCAAACTGAA 640
 QY 257 AspGlnArgProGlnLeuAspProTyGlnleuLeuGlyProThrSerSerArgLeuAla 276
 Db 641 GACCAAGCTCTCTAGTTAGATCTTATCAGATTTCTTGACCAACAGTAGCCCTTGCA 700
 QY 277 AsnProGlySerGlyGlnleuLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSer 296
 Db 701 AATCCAGCAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCGAGCTCTCTCGACAGC 760
 QY 297 SerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspPro 316
 Db 761 TCCAACTCCAGTGCATCAGCTGGGAGGACCAACCGGAGTTTCAAGATGACGGATCCC 820
 QY 317 AspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyAspLys 336
 Db 821 GACGAGGTGGCCGCGCTGGGAGAGCGGAAGAGCAACCCACATGAACTACGATAAG 880
 QY 337 LeuSerArgAlaLeuArgTyTyTyAspLysAsnleuMetThrLysValHisGlyLys 356
 Db 881 CTCACCGCGCCCTCCGTTACTACTATGACAAAGACATCATGACCAAGGTCCATGGGA 940
 QY 357 ArgTyAlaTyLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProPro 376
 Db 941 CGTAGCGCTCAAGTTCCAGCTTCCACGGGATCCCGCAGCCCTCCAGCCCAACCCCGC 1000
 QY 377 GluSerSerMetTyLysTyProSerAspLeuProTyMetSerSerTyHisAlaHis 396
 Db 1001 GAGTCATCTGTGTCAAGTACCCCTCGAGACCTCCCGTACATGGGCTCTATACGCCCCAC 1060
 QY 397 ProGlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSer 416
 Db 1061 CCACAGAGATGAATTTGTGGCGCCCACTCCAGCCCTCCCGTGACATTTCCAGT 1120
 QY 417 PhePheAlaAlaProAsnProTyTrpAsnSerProThrGlyGlyIleTyProAsnThr 436
 Db 1121 TTTTGTGTCGCCCAACCCATCTGGAATTCACCAACTGGGGTATATATACCCCAACT 1180
 QY 437 ArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyTyTy 451
 Db 1181 AGGCTCCCAACGACCATATGCTTCTCATCTGGGCACTTACTAC 1225

RESULT 8
 ID ABZ68769 standard; DNA; 1372 BP.
 AC ABZ68769;
 XX
 XX
 DT 16-MAY-2003 (first entry)
 XX
 DE Nucleotide sequence of human ERG splice variant C-1-4.
 XX
 KW Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
 KW C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
 KW cartilage disease; tumour; gene; ss.
 XX
 OS Homo sapiens.

XX Key Location/Qualifiers
 CDS 146..1309
 FT /*tag= a
 FT /product= "ERG splice variant C-1-4"
 XX
 PN WO2003000724-A2.
 PD 03-JAN-2003.
 XX
 XX 08-MAR-2002; 2002WO-EP002505.
 XX 25-JUN-2001; 2001US-0300756P.
 XX (TIGR-) TIGENIX NV.
 XX Luyten P, De Bari C, Dell'accio F;
 XX WPI; 2003-221398/21.
 XX P-PSDS; ABP97698.
 XX Novel splice variants of the Ets-related gene (ERG) transcription factor
 PT useful for preventing or treating bone-related or cartilage-related
 PT diseases or Ets-related gene associated tumor.
 XX
 PS Claim 112; Fig 7; 60pp; English.
 CC The present sequence encodes splice variant C-1-4 of human ERG. ERG
 CC belongs to the Ets family of transcription factors. The specification
 CC describes C-1-1 (characterised by the absence of the 81 bp fragment, and
 CC the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA
 CC sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp
 CC fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA
 CC sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and
 CC 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by
 CC the absence of the 218 bp fragment, and the presence of the 72 bp and 81
 CC bp fragments of the ERG-3 cDNA sequence). The splice variant
 CC polynucleotides are useful for stabilizing the phenotype of chondrocytes
 CC or cells with a chondrocytes-like phenotype, for preventing de-
 CC differentiation of articular chondrocytes during in vitro culture
 CC expansion, for preventing or treating bone-related or cartilage-related
 CC diseases or Ets-related gene associated tumour, or as marker of
 CC progenitor cells of stable chondrocytes. Cells expressing the
 CC polynucleotide are useful for producing or repairing cartilage tissue in
 CC a mammal
 XX
 SQ Sequence 1372 BP; 359 A; 430 C; 314 G; 269 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.02e-155 Length: 1372
 Score: 1926.00 Matches: 355
 Percent Similarity: 89.80% Conservative: 6
 Best Local Similarity: 88.31% Mismatches: 13
 Query Match: 78.81% Indels: 28
 DB: 8 Gaps: 2
 US-09-902-772-2 (1-451) x ABZ68769 (1-1372)
 QY 78 SerArgAsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSer 97
 Db 101 TCAAGGAACCTCTCTGATGAATGATGTCGCGCAAGAGCGGAGATGTTGGCGAGCCCA 160
 QY 98 AspAsnValGlyMetAsnTyTyTySerTyMetGluGluLysHisIleProProAsn 117
 Db 151 GACACCGTTGGATGAATCACTACGCGAGCTACATGAGGAGAGGACATGCCACCCCAAC 220
 QY 118 MetThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAsp 137
 Db 221 ATGACCAAGCAAGCAGCGCAGAGTTATCTGCCAGCAGATCTCTACGTTATGAGTACAGAC 280
 QY 138 HisValArgGlnTrpLeuGluTrpAlaValLysGluTyTyTyLeuProAspValAspIle 157
 Db 281 CATGTGGCGAGTGGCTGGAGTGGCGGTGAAGATATGSCCTTCCAGACGTCACATC 340

QY 158 LeuLeuPheGlnAsnIleAspGlyLysGluLeuCyseLysMetThrLysAspPheGln 177
Db 341 TTGTTATTCAGAACATCGATGGAGAACTGTGCAAGATGACCAAGACGACTTCAG 400
QY 178 ArgLeuThrProSerTyrsnAlaAspIleLeuLeuSerHisLeuHisTyrrLeuArgGlu 197
Db 401 AGGCTCACCCCAAGCTACACGCGGACATCCCTCTCTCACATCTCCACTACCTCAGAG 460
QY 198 Arg----- 198
Db 461 ACTCTCTTCCACATTTGACTTCAGATGATGTTGATAAGCCCTTACAAACCTCTCCACGG 520
QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrr 210
Db 521 TTAATGATCTAGAAACACAGCGGGGTGCGACTTTTATTTCCCAATATCTTCAGTATAT 580
QY 211 ProGluAlaThrGlnArgIleThrArgProAspLeuProTyrrGluGlnAlaArgArg 230
Db 581 CCTGAAGCTACGCAAGAATTAACAATAGGCGAGATTTACCATATAGCCGCCCCAGGAGA 640
QY 231 SerAlaThrThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSer 249
Db 641 TCAGCTTGACCGGTACGCGCCACCCACGCGCCAGTCGAAAGCTCTCAACATCTCT 700
QY 250 SerThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrrGlnIleLeuGly 269
Db 701 TCCACAGTGGCCAAACCTGAAGACGAGCGTCTCAGTTAGATCTTATCAGATTTCTTGA 760
QY 270 ProThrSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeu 289
Db 761 CCAACAAGTAGCCGCTTGCAATCCAGGAGTGCGGAGATCCAGCTTTGGCAGTTCTTC 820
QY 290 LeuGluLeuLeuSerAspSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGly 309
Db 821 CTGGAGCTCTGTCTGGACAGCTCCCACTCCAGCTGCATCCTCGGAAGGACCAACGGG 880
QY 310 GluPheLysMetThrAspProAspGluValAlaArgArgTrpGluArgLysSerLys 329
Db 881 GAGTTCAGATGACGGATCCGAGATCCGAGGTGGCGCGCTGGGGAGAGGAGAGCAAA 940
QY 330 ProAsnMetAsnTyrrAspLysLeuSerArgAlaLeuArgTyrrTyrrAspLysAsnIle 349
Db 941 CCCAACATGACTACGATAGCTACGCGCGCCCTCCGTTACTACTATGACAAAGACATC 1000
QY 350 MetThrLysValHisGlyLysArgTyrrAlaTyrrLysPheAspPheHisGlyIleAlaGln 369
Db 1001 ATGACCAAGTCTCATGGAGCGCTACGCTACAAAGTTCCACTTCCACGGGATCGCCAG 1060
QY 370 AlaLeuGlnProHisProGluSerSerMetTyrrLysTyrrProSerAspLeuProTyrr 389
Db 1061 GCCCTCCAGCCCCACCCCCCGAGTCTCTCTGACAGTACCCCTCAGACCTCCGCTAC 1120
QY 390 MetSerSerTyrrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAla 409
Db 1121 ATGGGCTCTATCAGCCGCCACCCACAGAAGATGAATTTGTGGCGCCGCCCTCCAGCC 1180
QY 410 LeuProValThrSerSerSerPheAlaAlaProAsnProTyrrTyrrAsnSerProThr 429
Db 1181 CTCGCCGGAGACTCTTCCAGTTTTTTCTGCTGCCCAACCCATCTGGAATTCACCACT 1240
QY 430 GlyGlyIleTyrrProAsnThrArgLeuProAlaHisMetProSerHisLeuGlyThr 449
Db 1241 GGGGGTATATACCCCAACACTAGGTCTCCACAGCATATGCCTTCTCTCTGGCACT 1300
QY 450 TyrrTyrr 451
Db 1301 TACTAC 1306
RESULT 9
ABZ68768
ID ABZ68768 standard; DNA; 1219 BP.
XX

AC ABZ68768;
XX 16-MAY-2003 (first entry)
DT Nucleotide sequence of human ERG splice variant C-1-3.
XX Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
XX C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
KW cartilage disease; tumour; gene; ss.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH 146..1156
FT CDS /*tag= a
FT /product= "ERG splice variant C-1-3"
XX WO2003000724-A2.
XX 03-JAN-2003.
XX 08-MAR-2002; 2002MO-EP002605.
XX 25-JUN-2001; 2001US-0300756P.
XX (TIGER-) TIGENIX NV.
PA Luyten F, De Bari C, Dell'accio F;
XX WPI; 2003-221398/21.
DR P-PSDB; ABP97697.
XX Novel splice variants of the Ets-related gene (ERG) transcription factor
PT , useful for preventing or treating bone-related or cartilage-related
PT diseases or Ets-related gene associated tumor.
XX Claim 11; Fig 6; 60pp; English.
XX The present sequence encodes splice variant C-1-3 of human ERG. ERG
CC belongs to the Ets family of transcription factors. The specification
CC describes C-1-1 (characterised by the absence of the 81 bp fragment, and
CC the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA
CC sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp
CC fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA
CC sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and
CC 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by
CC the absence of the 218 bp fragment, and the presence of the 72 bp and 81
CC bp fragments of the ERG-3 cDNA sequence). The splice variant
CC polynucleotides are useful for stabilizing the phenotype of chondrocytes
CC or cells with a chondrocytes-like phenotype, for preventing de-
CC differentiation of articular chondrocytes during in vitro culture
CC expansion, for preventing or treating bone-related or cartilage-related
CC diseases or Ets-related gene associated tumour, or as marker of
CC progenitor cells of stable chondrocytes. Cells expressing the
CC polynucleotide are useful for producing or repairing cartilage tissue in
CC a mammal
SQ Sequence 1219 BP; 312 A; 393 C; 288 G; 226 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 4,74e-145 Length: 1219
Score: 1807.50 Matches: 332
Percent Similarity: 90.40% Conservative: 7
Best Local Similarity: 88.53% Mismatches: 11
Query Match: 73.96% Indels: 25
DB: Gaps: 2
US-09-902-772-2 (1-451) x ABZ68768 (1-1219)
QY 78 SerArgAsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSer 97
Db 101 TCAAGGAAGTCTCTCTGATGATGCTGTCGCGGAGGAGATGTTGGCAGCCCA 160

```
Qy 98 AspAsnValGlyMetAsnTyrGlySerTyrMetGluGluHisIleProProProAsn 117
Db 161 GACACCGTTGGATGAATACGCGCAGCTACATGAGGAGGAGCAGCATGGCCACCCCAAC 220
Qy 118 MetThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTTPSerThrAsp 137
Db 221 ATGACACGAGACGAGCGAGAGTTATGTCGCCAGCAGATCTTACGCTATGGATGACAGAC 280
Qy 138 HisValArgGlnTrpLeuGluTrpAlaValIysGluTyrGlyLeuProAspValAspIle 157
Db 281 CATGTGGCGAGTGGCTGGAGTGGCGGTGAAGAATATGGCTTCCAGACGCTCAACATC 340
Qy 158 LeuLeuPheGlnAsnIleAspGlyIysGluLeuCysIysMetThrIysAspAspPheGln 177
Db 341 TTGTATTCCAGAACATCGATGGGAGGAACTGTGCAAGATGACCAAGGAGCACTTCCAG 400
Qy 178 ArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlu 197
Db 401 AGGCTCACCCCGCAGTCAACCGCGACATCTCTCTCACATCTCCACTACCTCAGAGAG 460
Qy 198 ArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIle 217
Db 460 ----- 460
Qy 218 ThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisser 237
Db 461 -----AATTACCATATGAGCCCGCCAGGAGATCAGCTGGACGGTCCAGGC 508
Qy 238 HisProThr-----GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGlu 256
Db 509 CACCCCGAGCCCGAGTGGAAAGCTGCTCAACCATCTCTTCCAGATGCCCAAACTGAA 568
Qy 257 AspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAla 276
Db 569 GACACAGCTCTCAGTTAGATCTTATCAGATCTTGGACCAACAGTACCGCGCTTGCA 628
Qy 277 AsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSer 296
Db 629 AATCCAGGACGTGGCCAGATCCAGCTTGGCAGTCTCTCTGGAGCTCCTGTGGACAGC 688
Qy 297 SerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspPro 316
Db 689 TCCAACTCCAGCTGCATCACCCTGGAGGACCAACCGGGAGTTTCAGATGACGGATCCC 748
Qy 317 AspGluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLys 336
Db 749 GACGAGGTGGCGCGCTGGGAGAGCGGAGAGCAAAACCAACATGAACTACGATAAG 808
Qy 337 LeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLys 356
Db 809 CTCAGCCGCGCCCTCGGTACTACTATGACAGAAACATCATGACCAAGGTCCATGGGAAG 868
Qy 357 ArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProPro 376
Db 869 CGCTACGCTCAAGTTCGATCTTCCAGGGATGCGCCAGCGCTCCAGCCCGCCCGCCCG 928
Qy 377 GluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHis 396
Db 929 GAGTCATCTGTACAGTACCCCTCAGACCTCCCGTACATGGGCTCTCTATCAGCGCCAC 988
Qy 397 ProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSer 416
Db 989 CCACAGAAAGTGAATTTGTGGCGCCCGCCCGCTCCAGCCCTCCCGGTGACATCTTCCAGT 1048
Qy 417 PhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyIleTyrProAsnThr 436
Db 1049 TTTTGTGTCGCCCAACCCCATCTGGAATTCACCACTGGGGGTATATACCCCAACACT 1108
Qy 437 ArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1109 AGGCTCCCGACAGCCATATGCTTCTCATCTGGGCATTTACTAC 1153
```

RESULT 10

```
ADA02692
ID ADA02692 standard; cDNA; 1359 BP.
XX
AC ADA02692;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human FLI1 carcinoma associated coding sequence, SEQ ID NO:1210.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
(SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
XX useful for preparing compositions for treating carcinomas.
XX
PS Claim 1; SEQ ID NO 1210; 245pp; English.
XX
CC The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1359 BP; 359 A; 422 C; 333 G; 245 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 2,52e-129 Length: 1359
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
DB: Gaps: 7
```

US-09-902-772-2 (1-451) x ADA02692 (1-1359)

```
Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
Db 1 ATGACGGGACATATTAAAGAGGCTCTGTCGGTGGGAGCAGCACCGATCCCTCTTTGAC 60
```

Qy 21 CysAlaTyrGly---SerProHisLeuAlaIysThrGluMetThrAlaSerSerSer 39
 Db 61 TCAGCGTACGAGCGGAGCCCATCTCCCAAGAGCCGACATGACTGCTCGGGAGTCCT 120
 Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
 Db 121 GACTACGGGAGCCCAACAGATCAACCCCTCCACACAGAGGATGATCAATCAG 180
 Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
 Db 181 ---CCAGTGGGGTCAACGTCAGCGGAGTAT-----GACCACATGAATGGATCCAGG 231
 Qy 80 AsnSerProAspAspCysSerValAlaIysGlyLysMetValSerSerSerAspAsn 99
 Db 232 GAGTCTCGGTGACTGACGGTTAGCAATGACAGAGCTGGTGGCGAGCGGAGTCC 291
 Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118
 Db 292 AACCCCATGAACACTACACAGCTATATGAGCAGAGAAATGGCCCCCTCTCCCAACATG 351
 Qy 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
 Db 352 ACCACCAACGAGAGAGAGTATCGTCCCGCAGACCCACACATGTGGACACAGGAGCAT 411
 Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
 Db 412 GTGAGGCAATGGCTGGAGTGGCCCATAAAGAGGTACAGCTTGATGGAGATCGACACATCC 471
 Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
 Db 472 TTTTTCAGAAACATGGATGGCAAGACTGTGTAAATGAACAGGAGGACTTCTCTCCGC 531
 Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
 Db 532 GCCACCCCTCTACACACGGAAGTGTGTGTGCACACCTCAGTTACTCTCAGGAA-- 588
 Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
 Db 589 ---AGTTCACGTGCTGGCTATTAACAACTCCACACCGCAACCTCTCAGCATTTGAGT 645
 Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHis 238
 Db 646 GTCAAGAGAGACCTCTTATGCTCAGTCAGTACAGAGAGGAGCTTGGGCAATACATGAT 705
 Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
 Db 706 TCTGCTCCAAACAAAGTCTCCCTTGGAGGGGCGACAAACGATCAGTAAGAATACAGAG 765
 Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
 Db 766 CACGGCCCCAGCCAGATCCGTATCAGATCCTGGGCGGACACAGCATGCTAGCCACAC 825
 Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSer 297
 Db 826 CTTGAAGCGGGCAGATCCAGCTGTGGCAATTCCTCTGGAGCTGCTCCGACAGGGCC 885
 Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
 Db 886 AACGCCACTGTATCCTCTGGAGGGGACCAACGGGGAGTTCAAAATGACGGACCCCGAT 945
 Qy 318 GluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
 Db 946 GAGGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGAATTAACGACAGCTG 1005
 Qy 338 SerArgAlaLeuArgTyrTyrThrLysAsnIleMetThrLysValHisGlyLysArg 357
 Db 1006 AGCCGGGCGCTCGTTATCTATATATAAAACATATGACCAAGGTGACGGCAAGA 1065
 Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
 Db 1066 TATGTTTACAAATTTGATCTCCAGCGCATTCGCCAGGCTCTGCAGCCACATCCGACCGAG 1125
 Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397

Db 1126 TCGTCCATGTACAGTACCCCTTCGACATCTCTCATGCTCTTCTACATGCCACAG 1185
 Qy 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
 Db 1186 CAGAAAGGTGAACCTTGTCCCTCCCATCCATCTCCATGCTGTCTCTCTCCAGCTTC 1245
 Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435
 Db 1246 TTTGAGCGGCATCAATACTGGACCTCCCGCGGGGAATCTACCCACCCCAAC 1305
 Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1306 GTCCCGCCCATCTCTTACACCCACGTCGCTTACACATTAGCAGCTACTAC 1356

RESULT 11

ADB72430

ID ADB72430 standard; cDNA; 1359 BP.

XX ADB72430;

DT 04-DEC-2003 (first entry)

DE Human FLI1 cDNA.

XX human; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
 KW cancer; neoplasm; adenocarcinoma; sarcoma; Gene.
 XX Homo sapiens.

XX WO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-0004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,

XX cancers, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 258; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the 660 sequences fully defined
 CC in the specification. A polynucleotide of the invention has cytostatic
 CC activity, and may have a use in gene therapy, or in a vaccine. The
 CC recombinant nucleic acids and polypeptides are useful for treating
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
 CC sarcomas. The present sequence represents a human cDNA of the invention.

XX Sequence 1359 BP; 359 A; 422 C; 333 G; 245 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2,52e-129	Length:	1359
Score:	1624.50	Matches:	306
Percent Similarity:	79.43%	Conservative:	57
Best Local Similarity:	66.96%	Mismatches:	83
Query Match:	9	Indels:	11
DB:		Gaps:	7

US-09-902-772-2 (1-451) x ADB72430 (1-1359)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValIserGluAspGlnSerLeuPheGlu 20
 Db 1 ATGACGGGACTATTAAGGAGGCTGTGGTGGTGGAGCAGCAGCCCTCTTTGAC 60
 Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
 Db 61 TCAGCGTACCGAGCGGAGCCCTCTCCCAAGCGGACATGATCGCTCGGGAGTCT 120
 Qy 40 GlnTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
 Db 121 GACTACGGGAGCCGCCACAGATCAACCCCTCCACACAGCAGGAGTGGATCAATCAG 180
 Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
 Db 181 ---CCAGTGGGTCAACGTCACAGCGGGAGTAT-----GACCACATGAATGATCCAG 231
 Qy 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAsp 99
 Db 232 GAGTCTCCGTGAGCTGACGCTTAGCNAATGACGACAGCTGGTGGCGGAGCGAGTCC 291
 Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118
 Db 292 AACCCCATGAATACACAGCTATATGGACGAGAAGATGGCCCCCTCTCCCAACATG 351
 Qy 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
 Db 352 ACCACCAACAGAGGAGAGTCACTGTCGCCCGACGCCACACTGTGGACACAGGAGCAT 411
 Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
 Db 412 GTGAGGCAATGGCTGGAGTGGCCCATAAAGAGTACAGCTTGATGGAGATCGACATCC 471
 Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
 Db 472 TTTTTCAGAACATGATGGTCAAGAGCACTGTGTAAATGAACAAGGAGCACTCTCCGC 531
 Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
 Db 532 GCACACACCTCTACAAACGAGAGTGTCTGTACACCTCTGATCTCTACAGGNA--- 588
 Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
 Db 589 ---AGTTCACTGTGGCTATAATACAACTCCACACGACCACTCTCAGCATGAGT 645
 Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHis 238
 Db 646 GTCAAGAAGACCCCTCTTATGACTCAGTCAGAAGAGAGCTTGGGGCAATACATGAAT 705
 Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
 Db 706 TCTGGCTCAACAAAGTCTCCCTTGGAGGGGCACAAACGATCAGTAAGAATACAGAG 765
 Qy 258 GluArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
 Db 766 CAACGGCCCCCAGCAGATCGGTATCAGATCTCGGGCCGACACGAGTCCGCTAGCAAC 825
 Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuSerSerSerSer 297
 Db 826 CTTGGAAAGCGGCGAGATCCAGCTGTGGCAATCTCTCGAGGTGCTCTCCGACAGCGCC 885
 Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
 Db 886 AACGCCAGCTGTATCACTGGGAGGGGACCAACGGGGAGTTCAAAATGACGGACCCCGAT 945
 Qy 318 GluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
 Db 946 GAGTGGCCAGGCGCTGGGCGGAGCGGAAAGCAAGCCCAACATGATTAACGACAGCTG 1005
 Qy 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
 Db 1006 AGCCGGGCCCTCCGTATTACTATGATAAAACATTATGACCAAGAGTGCAGGCAAAAGA 1065
 Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu 377

Db 1066 TATGCTTACAAATTTGACTTCACGGCATTCGCCAGGCTCTGACGCCACATCCAGCCAG 1125
 Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
 Db 1126 TCGTCCATGTACAAGTACCTTCTGACATCTCTACATGCTCTCTACATGCCACCCAG 1185
 Qy 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
 Db 1186 CAGAAGGTGAACCTTTGCTCCCTCCCATCCATCCATCCATCCATCCATCCATCCATCC 1245
 Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435
 Db 1246 TTTGAGCGCGCATCACATCTGACCTTCCCTCCACGGGGGAATCTACCCCAACCCCAAC 1305
 Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1306 GTCCCCCGCCATCTAACACCCAGCTGCCITTCACATAGGACGCTACTAC 1356
 RESULT 12
 AAQ50644
 ID AAQ50644 standard; cDNA; 2938 BP.
 XX AC AAQ50644;
 XX DT 25-MAR-2003 (revised)
 DT 26-MAY-1994 (first entry)
 XX DE Human Hum-Flt-1 gene clone BM025.
 XX KW chromosomal translocation; chimeric; chimeraic; Ewing sarcoma; Ews gene;
 KW malignant melanoma; hum-flt-1;
 KW primitive peripheral neuroectodermal tumour; human chromosome 11;
 XX human chromosome 22; ss.
 OS Homo sapiens.
 XX FH Key
 FT CDS Location/Qualifiers
 FT 143..1501
 FT /*tag= a
 FT polyA_signal /product= "HUM-FLI-1"
 FT 2908..2913
 FT /*tag= b
 XX PN W09323549-A2.
 XX PD 25-NOV-1993.
 XX PF 19-MAY-1993; 93WO-FR000494.
 XX PR 20-MAY-1992; 92FR-00006123.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX PA Aurias A, Delattre O, Desmase C, Melot T, Peter M, Plougastel B;
 PI Thomas G, Zucman J;
 XX DR P-PSDB; AAR44556.
 XX WPI: 1993-386580/48.
 XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence
 PT involved in chromosomal translocation, also derived mRNA, probes, fusion
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
 XX Disclosure; Fig 7; 123pp; French.
 XX PS The probe 11RR1 was used to screen a human marrow cDNA library (Clontech
 CC cat.# HL1058). The clone EM025 was identified and sequenced. It
 CC represents the entire coding region together with 5'- and 3'-UTRs of the
 CC Hum-Flt-1 gene. (Updated on 25-MAR-2003 to correct PN field.)
 XX Sequence 2938 BP; 847 A; 692 C; 675 G; 724 T; 0 U; 0 Other;

Alignment Scores:

```

Pred. No.: 7,22e-129 Length: 2938
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
DB: 2 Gaps: 7

US-09-902-772-2 (1-451) x AAQ50644 (1-2938)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 143 ATGACGGGAGTATTAGAGGCTCTGTCGTGTGAGCAGCAGCAGCAGCCTCTTGTGAC 202
Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 203 TCAGCGTACGAGCGGCGAGCCATCTCCCAAGCGCGACATGACTCGCTCGGGAGTCT 262
Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
Db 263 GACTACGGGAGCGGCCCAAGATCAACCCCTCCACACAGCAGGAGTGGATCAATCAG 322
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db 323 ---CCAGTGGGTCAACGTCAGCGGAGTAT-----GACCACATGATGGATCCAGG 373
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
Db 374 GAGTCTCCGTGGACTCAGCTAGCAATGACGCAAGCTGTGGCGGAGCGGAGTCC 433
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118
Db 434 AACCCTGATGACTACACACTATATGGACGAGAGATGGCCCCCTCTCTCCACATG 493
Qy 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 494 ACCACCAACGAGAGGAGTCACTGCTCCCGCAGACCCACACATGTGGACACAGGAGCAT 553
Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
Db 554 GTGAGGCAATGCTGGAGTGGGCCATTAAGAGTATAGCTTGATGAGATCGACATCC 613
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 614 TTTTTCAGAACATGATGCAAGGAATCTGTAAATGAACAAGGAGGACTTCTCCCGC 673
Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
Db 674 GCCACCACTCTACACACGGAAGTCTGTGTACACCTCAGTTACCTCAGGAA--- 730
Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db 731 ---AGTTCACCTGCTGGCTATATACAACTCCACACCGACCAATCTCCAGGATGAGT 787
Qy 219 ThrArgProAspLeuProTyrGluAlaArgSerAlaTrpThrSerHisSerHis 238
Db 788 GTCAAGAGAGACCTCTTATGACTCAGTCAGAGAGAGCATGTGGGCAATAACATGAAT 847
Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
Db 848 TCTGGCTCAACAAGAGTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 907
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 908 CAACGGCCCCAGCAGATCCGTATCAGATCTCGGGCCCGACGACGATCGCTAGCAAC 967
Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeuLeuSerSerSer 297
Db 968 CTTGAGAGCGGCGAGATCCAGCTGTGGCAATTCCTCTGAGTGTCTCTCCGACGGGCC 1027
Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 1028 AACGCCAGCTGTATCCTCGGAGGGGACCAACGCGGAGTTCAAAATGACGGACCCCGAT 1087

```

```

Qy 318 GluValAlaArgArgTTPGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 1088 GAGGTGGCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGAATACGACAGCTG 1147
Qy 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1148 AGCGGGCCCTCCGTATTACTATGATAAAACACATTATGACCAAGTGCAGCGCAAGA 1207
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1208 TATGCTTACAAATTGACTTCCAGCGCATTCGCCAGGCTCTGCAGCCACATCCGACCG 1267
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1268 TCGTCCATGTACAAGTACCTTCTGACATCTCTTACATGCTTCTCTACCATGCCCCAC 1327
Qy 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
Db 1328 CAGAAGTGAATTTGTCCTCCCTCCATCCATCTCCATGCTGTCACTTCTCTCCAGCTTC 1387
Qy 418 PheAlaAlaProAsnProTyrTyrTyrAsnSerProThrGlyGlyIleTyrProAsn---- 435
Db 1388 TTTGGAGCGCGCATCACAATACTGGACCTCCGCCACGGGGGATCTACCCCAACCCCAAC 1447
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1448 GTCCCCCGCATCTCTAACACCCCGTGCCTTCACACTTAGCAGCTACTAC 1498

RESULT 13
ABV94304
ID ABV94304 standard; cDNA; 2957 BP.
AC ABV94304;
XX
DT 08-JAN-2003 (first entry)
XX
DE Breast carcinoma related nucleotide sequence SEQ ID NO:295.
XX
KW Human; breast carcinoma; cancer; tumour; cytostatic; anti-tumour; gene;
XX
OS Homo sapiens.
XX
PN WO200246467-A2.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001WO-IB002811.
XX
PR 08-DEC-2000; 2000US-0254090P.
XX
PR 07-DEC-2001; 2001US-00007926.
XX
PA (IPSO-) IPSOGEN.
XX
PI Bertucci F, Houlgatte R, Birnbaum D, Nguyen C, Viens P, Fert V;
XX
DR WPI; 2002-619023/66.
XX
PT Novel polynucleotide library useful in molecular characterization of a
PT carcinoma, comprising a pool of polynucleotide sequences or its
PT subsequence which are either underexpressed or overexpressed in tumor
PT cells.
XX
PS Claim 1; Page 300-301; 401pp; English.
XX
CC The present invention describes a polynucleotide library (I) useful in
CC the molecular characterisation of a carcinoma, comprising a pool of
CC polynucleotides or its subsequences which are either underexpressed or
CC overexpressed in tumour cells, and correspond to any of the
CC polynucleotide sequences chosen from the 488 sequences given in ABV9401.0
CC to ABV94477. Also described: (i) a polynucleotide array (II) useful for
CC the prognosis or diagnostic of tumour, comprising (I); and (2) detecting
CC

```

(M1) differentially expressed polynucleotide sequences which are correlated with a cancer, involves obtaining a polynucleotide sample from a patient, and reacting the polynucleotide sample obtained with a probe immobilised on a solid support, where the probe comprises any combination of the polynucleotide sequences of (I) or its expression products encoded by polynucleotide sequences of (I), and detecting the reaction product. (I) have cytostatic activities and can be used as anti-tumour agents. (I) is useful for the prognosis or diagnosis of a carcinoma. (I) and (II) are useful for the prognosis or diagnosis of tumour, in differentiating a normal cell from a cancer cell, detecting a hormone sensitive tumour cell, differentiating a tumour with lymph nodes from a tumour without lymph nodes, differentiating antracycline-sensitive tumours from antracycline-insensitive tumours, and classifying good and poor prognosis primary breast tumours. (I) is useful for large-scale molecular characterisation of breast cancer that help in prediction, prognosis and cancer treatment, and for detecting differentially expressed genes that correlated with a cancer

XX Sequence 2957 BP; 855 A; 698 C; 680 G; 724 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,28e-129 Length: 2957
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
DB: 6 Gaps: 7

US-09-902-772-2 (1-451) x ABV94304 (1-2957)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 173 ATGACGGGACTATTAGAGGGCTCTGCGTGGTGGAGCAGCACCAGTCCCTTTTGAC 232
Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 233 TCAGCGTACGGAGCGGAGCCCATCTCCCAAGCCGACATGACTGCTCGGGAGTCTT 292
Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
Db 293 GACTACGGGAGCCGCCACAGATCAACCCCTCCACAGCAGGAGGATCAATCAG 352
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValaAnglySerArg 79
Db 353 ---CCAGTGGGGTCAACGTCAAGCGGAGTAT-----GACCACATGAATGATCCAGG 403
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99
Db 404 GAGTCTCCGGTGACTCAGCGGTAGCAAAATGAGCAAGCTGTGGCGGAGCGAGTCC 463
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118
Db 464 AACCCCATGAATACACAGCTATATGGACGAGAGAAATGGCCCCCTCTCCCAACATG 523
Qy 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 524 ACCACCAACGAGAGGAGTATCGTCCCGCAGACCCCACTGTGGACACAGGAGCAT 583
Qy 139 ValArgGlnTrpLeuGluTTPAlaValLysGluTyrGlyLeuProAspValaAspIleLeu 158
Db 584 GTGAGGCAATGGCTGGAGTGGCCATAAAGAGTACAGTTGATGGAGATCGACATCC 643
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysGlyLysMetThrLysAspAspPheGlnArg 178
Db 644 TTTTTCAGAACATGATGGTGAAGAACTGTATAAATGAACAGGAGGACTTCTCCCGC 703
Qy 179 LeuThrProSerTyrAnAlaAspIleLeuLeuSerHisIleHisTyrLeuArgGluArg 198
Db 704 GCACCAACCTCTACAAACAGGAGTCTGTGTACACCTCAGTACCTCAGGAA--- 760
Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db 761 ---AGTTCACTGCTGGCCTATAATACAACTCCACACCGCACCAATCCTCACGATGAGT 817

Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTTPThrSerHisSerHis 238
Db 818 GTCAAGAGAGCCCTTCTATGACTCACTCAGAGAGAGCTTGGGCAATACATAGAT 877
Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
Db 878 TCTGGCTTCAACAAAGTCTCTCCCTTGGAGGGGCACAAAGATCAGTAAGAAATACAG 937
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 938 CAACGGCCCCAGCAGATCCGATCAGATCTTGGCCCCGACGAGTGCCTAGCCAC 997
Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerSerSer 297
Db 998 CCTGGAAGCGGCAGATCAGCTGTGGCAATTCCTCTCTGGAGCTCTCTCCGACAGCC 1057
Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 1058 AACGCCAGCTGTATCACCCTGGAGGGGACCAACGGGGAGTTCAAAATGACGAGCCCGAT 1117
Qy 318 GluValAlaArgArgTTPGlyLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 1118 GAGGTGGCCAGCGCTGGGGCGAGCGGAAAGCAAGCCCAACATGAATTACGACAAGCTG 1177
Qy 338 SerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1178 AGCCGGGCCCTCCGTTATTACTATGATAAAACATTTATGACCAAGTGCACGGCAAGA 1237
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1238 TATGCTTACAAATTTGACTTCCAGGCATTGCCAGGCTCTGCAGCCACATCCGACCGAG 1297
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1298 TCGTCCATGTACAAGTACCTCTGACATCTCTACATGCTCTCTACATGCCCCACCG 1357
Qy 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417
Db 1358 CAGAAGGTGAATTTGTCTCCCTCCCATCTCTCCATCTCTGCTGCTCTCTCCAGCTTC 1417
Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db 1418 TTTGGAGCGCATCAACAATACATACTGGACCTCCCGCCAGCGGGGAATCTACCCCAACCCCAAC 1477
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1478 GTCCCCCGCCATCTTAACACCCAGCTGCTTACACATTTAGGACGCTACTAC 1528
RESULT 14
ABK84139
ID ABK84139 standard; cDNA; 2957 BP.
XX ABK84139;
AC ABK84139;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #710.
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
OS Homo sapiens.
XX
XX WO200228999-A2.
XX
XX
PD 11-APR-2002.

XX 03-OCT-2001; 2001WO-US030821.
XX 03-OCT-2000; 2000US-0237189P.
XX (GENE-) GENE LOGIC INC.
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX Detecting granulocyte activation by detecting differential expression of
XX genes associated with granulocyte activation, which serves as diagnostic
XX markers that is useful for monitoring disease states and drug toxicity.
XX Claim 1; SEQ ID NO 710; 114pp; English.
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing the
XX expression level to an expression level in an unactivated GC, where
XX differential expression of Gs is indicative of GCA. Also included are
XX modulating (M2) GCA by contacting GC with an agent that alters the
XX expression of at least one gene in Gs; (2) screening (M3) for an agent
XX capable of modulating GCA or an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease using the gene expression
XX profile; (3) detecting (M4) an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease, by detecting the level of
XX expression in a sample of the tissue of gene(s) from Gs, where the level
XX of expression of the gene is indicative of inflammation; (4) treating
XX (M5) an inflammation (especially chronic) or in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease, by contacting a tissue having inflammation with an
XX agent that modulates the expression of gene(s) from Gs in the tissue. M1
XX is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful
XX for screening an agent capable of modulating GCA preferably in an
XX inflammation in a tissue; M4 is useful for detecting an inflammation
XX (especially chronic) in a tissue, an allergic response in a subject,
XX exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
XX psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
XX cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
XX respiratory distress syndrome, inflammatory bowel disease, Crohn's
XX disease, ulcerative colitis, periodontal disease; also bacterial
XX infection, viral infection, parasitic infection, protozoal infection,
XX fungal infection and M5 is useful for treating one of the above
XX conditions. The present sequence represents a gene differentially
XX expressed in granulocytes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pat_sequences
XX
XX SQ Sequence 2957 BP; 855 A; 698 C; 680 G; 724 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 7,28e-129 Length: 2957
XX Scores: 1624.50 Matches: 306
XX Percent Similarity: 79.43% Conservative: 57
XX Best Local Similarity: 66.96% Mismatches: 83
XX Query Match: 66.47% Indels: 11
XX DB: 6 Gaps: 7
XX
XX US-09-902-772-2 (1-451) x ABK84139 (1-2957)
XX
XX QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
XX DB 173 ATGGACGGGACTATTAAAGAGGCTCTGTCGGTGTGGAGCGACGACCATCTTTGAC 232
XX QY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
XX DB 233 TCAGGGTACGGAGCGGCGACGCCATCTCCCAAGGCCGACATGATGCTCGGGAGTCTCT 292

QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
DB 293 GACTACGGGCGAGCCCAACAGATCAACCCCTCCACACAGCAGGAGTGGATCAATCAG 352
QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
DB 353 ---CCAGTGAGGGTCAACGTCAAGCGGAGTAT-----GACCACATGAATGGATCCAGG 403
QY 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99
DB 404 GAGTCTCCGGTGGAGCTGCAGCGTTAGCAATGTCAGCAAGCTGGTGGCGGAGGAGTCC 463
QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGlyLysHis---IleProProProAsnMet 118
DB 464 AACCCCATGAACACTCAACAGCTATATGACGAGAGAAGTGGCCCTCTCTCCCAACATG 523
QY 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
DB 524 ACCACCAACGAGAGAGAGTATCTGTCCTCCGACACCCACACTGTGGACAGAGGAT 583
QY 139 ValArgGlnTrpLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
DB 584 GTGAGCATGGCTGGAGTGGCCATAAAGGAGTACAGCTTGATGGAGATCGACATCC 643
QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
DB 644 TTTTTCAGAACATGGATGGCAAGCACTGTGTAATGAACAAGGAGGACTTCTCCGCG 703
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
DB 704 GCCACACCTCTACACACGGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 760
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
DB 761 ---AGTTCACCTGCTGGCTCTATATACAACTCCCAACGACCAATCTCAGATTTAGT 817
QY 219 ThrArgProAspLeuProTyrGlnAlaArgSerAlaTyrThrSerHisSerHis 238
DB 818 GTCAAGAAGACCTCTTATGATCTCAGTCAAGAGAGAGCTTGGGCAATAACATGAT 877
QY 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
DB 878 TCTGGCTCTCAACAAAGTCTCTCTCTGGAGGGGCAACAGCATCAGTAAAGTAAACAGAG 937
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
DB 938 CAACGGCCCCGACGAGATCCGATCTGTCGGCCCGACACGAGCTCGCTAGCCAC 997
QY 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerSerSer 297
DB 998 CTTGGAAGCGGCGACATCCAGCTGTGGCAATCTCTCTGGAGTGTCTCTCCGACAGCGCC 1057
QY 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
DB 1058 AACGGCAGCTGATCACCTGGAGGGGACCAACGGGAGTTCAAAATACGGACCCCCCAT 1117
QY 318 GluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
DB 1118 GAGGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCAACATGAATAGTACAGAGCTG 1177
QY 338 SerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLysValHisGlyLysArg 357
DB 1178 AGCCGGGGCCCTCCGTTATTACTATGATAAAACATTTAGCAAAAGTGCAGCGCAAGA 1237
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu 377
DB 1238 TATGCTTACAAATTGACTTCCACGCGATTCGCCAGGCTCTGCAGCCACATCCGACCGAG 1297
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
DB 1298 TCGTCCATGATACAGTACCCCTCTGACATCTCTGATGCTCTCTGATGCTCTCTGATGCT 1357
QY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417

Db	1358	CAGAAGTGTGAACCTTGTGTCCTCCCTCCCATCCATCTCCATGCTGTCACTTCTCCAGTTC	1417
Qy	418	PheAlaAlaProAsnProTyrTTPAsnSerProThrGlyGlyIleTyrPProAsn-----	435
Db	1418	TTTGAGCCGGCATCACAAATCTGACACCTCCCCACGGGGGAATCTACCCACCCCAAC	1477
Qy	436	---ThrArgIeuProAlaAlaHisMetProSerHisIeuGlyThrTyrTyr	451
Db	1478	GTCGCCCGCCATCTCTAACCCACGCTGCTTTCACACTTAGGCAGCTACTAC	1528
RESULT	15		
ADAO2691			
ID	ADA02691	standard; cDNA; 2957 BP.	
AC	ADA02691;		
CC	XX		
DD	XX	06-NOV-2003 (first entry)	
DE	XX	Human FLI1 carcinoma associated cDNA, SEQ ID NO:1209.	
EE	XX		
KW	XX	Human; carcinoma associated; oncogene; carcinoma; cancer; breast;	
KW	XX	prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;	
KW	XX	gene; ss.	
OS	XX	Homo sapiens.	
PN	XX	WO2003057146-A2.	
PD	XX	17-JUL-2003.	
PF	XX	26-DEC-2002; 2002WO-US041414.	
PR	XX	26-DEC-2001; 2001US-00035832.	
PA	XX	(SAGR-) SAGRES DISCOVERY.	
PI	XX	Morris DW;	
DR	XX	WPI; 2003-587068/55.	
PT	XX	New recombinant nucleic acid encoding carcinoma associated protein,	
PT	XX	useful for preparing compositions for treating carcinomas.	
PS	XX	Claim 1; SEQ ID NO 1209; 245pp; English.	
CC	XX	The invention relates to recombinant carcinoma associated (CA) nucleic	
CC	XX	acid sequences from mouse and human (ADA01482-ADA03094), and to	
CC	XX	recombinant carcinoma associated proteins (CAP) encoded by them. The	
CC	XX	invention also encompasses expression vectors and host cells comprising a	
CC	XX	CA nucleic acid, a polypeptide (especially an antibody) that specifically	
CC	XX	binds to the protein, and a bioclip comprising CA nucleic acid or	
CC	XX	fragments thereof. The sequences of the invention were identified using	
CC	XX	oncogenic retroviruses, which insert into the genome of the host organism	
CC	XX	at random. Many of these do not carry transduced host oncogenes or	
CC	XX	pathogenic trans-acting viral genes, meaning that cancer incidence is a	
CC	XX	direct consequence of the effects of proviral integration into host	
CC	XX	protooncogenes. The CA nucleic acid sequences can be used to diagnose	
CC	XX	carcinoma (especially breast cancer, prostate cancer, lymphoma or	
CC	XX	leukaemia) or a propensity to carcinoma by determination of the sequence	
CC	XX	of a CA gene, or by determination of CA gene expression in particular	
CC	XX	tissues. CA nucleic acids, proteins and antibodies are also useful as	
CC	XX	therapeutic agents and in screening and evaluating drug candidates. The	
CC	XX	present sequence represents a specifically claimed human CA nucleic acid	
CC	XX	sequence of the invention. Note: The complete sequence data for this	
CC	XX	patent did not form part of the printed specification, but was obtained	
CC	XX	in electronic format directly from WIPO at	
CC	XX	ftp.wipo.int/pub/published_pct_sequences.	
XX	XX	Sequence 2957 BP; 855 A; 698 C; 680 G; 724 T; 0 U; 0 Other;	

Alignment Scores:		
Pred. No.:	7.28e-129	Length: 2957

Score:	1624.50	Matches:	306
Percent Similarity:	79.43%	Conservative:	57
Best Local Similarity:	66.96%	Mismatches:	83
Query Match:	66.47%	Indels:	11
DB:	8	Gaps:	7

US-09-902-772-2 (1-451) x ADA02691 (1-2957)

Qy	1	MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu	20
Db	173	ATCGAGCGGACTATTAAAGAGGCTCTGCTGGTGGTGGAGCGACACCAAGTCCCTCTTTGAC	232
Qy	21	CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer	39
Db	233	TCACGCTACGAGCGGACGCCCATCTCCCAAGGCCGACATGACTGCTCGGGAGTCCCT	292
Qy	40	GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln	59
Db	293	GACTAGCGGCGAGCCCCACACAGATCAACCCCTCCACACAGCAGGAGTGGATCAATCAG	352
Qy	60	ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg	79
Db	353	---CCAGTGAGGGTCAACGTCAAGCGGGAGTAT-----GACCACATGAATGATCCAGG	403
Qy	80	AsnSerProAspAspCysSerValAlaLysGlyGlyMetValSerSerSerAspAsn	99
Db	404	GAGTCTCCGGTGACTGCAGCGTTAGCAAAATGAGCAAGCTGGTGGCGGAGCGAGTCC	463
Qy	100	ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProProAsnMet	118
Db	464	AAOCCCATGAACATACACAGCTATATGGACGAGAAGAAATGGCCCCCTCTCTCCCAATG	523
Qy	119	ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis	138
Db	524	ACACCAACAGAGGAGAGTCACTGTCCTCCGACAGACCCACACTGTGGACAGGAGCAT	583
Qy	139	ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu	158
Db	584	GTGAGCAATPGTGGAGTGGGCATAAAGAGGTACAGCTTGATGGAGATCGACATCC	643
Qy	159	LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg	178
Db	644	TTTTTCCAGAACATGATGTCGACAGAACTGTGTAAATGAAACAGGAGGACTCTCTCCGC	703
Qy	179	LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg	198
Db	704	GCCACCAACCCTTACAAACACGGAAGTGTGTTGTCAACCTCAGTTACCTCAGCGGAA--	760
Qy	199	GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr	218
Db	761	---AGTTCACTGTGGCCTATAATACAACTCCACACCGACCAATCTCTCAGGATTGAGT	817
Qy	219	ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHis	238
Db	818	GTCAAGAGACCCCTTCTTATGACTCAGTCAGAAGAGGAGCTTTGGGCAATAACATGAAT	877
Qy	239	ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp	257
Db	878	TCCTGGCCTCAACAAAAGTCTCCCTCTGGAGGGGCAACACGATCAGTAAGAATACAGAG	937
Qy	258	GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn	277
Db	938	CAACGGCCCGCAGCATCGTATCAGATCTCTGGGCCCGACACAGCAGTCCCTAGCCAC	997
Qy	278	ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluIleuSerAspSerSer	297
Db	998	CCTGGAAGCGGCGAGATCCAGCTGTGGCAATCTCTCTCTGGAGTGTCTCTCCGACAGGCC	1057
Qy	298	AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp	317
Db	1058	AACGCCAGCTGTATCACTCTGGAGGGGACCAACGGGGAGTTCAAAATGACGGACCCCGAT	1117
Qy	318	GluValAlaArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu	337

```
Db      1118 GAGGTGGCCAGCGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGAAATACGACAAAGCTG 1177
Qy      338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db      1178 AGCCGGGCGCTCCGTTATTACTATGATAAAAAACATTATGACCAAAAGTGACGCGCAAAAGA 1237
Qy      358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db      1238 TATGCTTACAAATTGACTTCCAGCGCATTTGCCAGGCTGTGCAGCCACATCCGACCGAG 1297
Qy      378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerTyrHisAlaHisPro 397
Db      1298 TCGTCCATGTACAGTACCTTCTGACATCTCTTACATGCTTCTTACCATGCCACCAG 1357
Qy      398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417
Db      1358 CAGAAGGTGAACCTTTGTCCCTCCCATCCATCCCTCCATGCTGTCACCTTCTCCAGCTTC 1417
Qy      418 PheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db      1418 TTTGAGCGCGCATCACAATATCTGGACCTCCCCACGGGGGGAATCTACCCCAACCCCAAC 1477
Qy      436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db      1478 GTCCCCCGCCATCCTTAACACCCACGTCGCTTACACACTTAGGCAGCTACTAC 1528
```

Search completed: May 27, 2004, 21:22:13
Job time : 642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 27, 2004, 20:59:17 ; Search time 128 Seconds
(without alignments)
1955.335 Million cell updates/sec

Title: US-09-902-772-2
Perfect score: 2444
Sequence: 1 MASTIKALSVSDQSLFE.....LYPNTRLPAAHMPSHLGTYV 451

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US0902772/runat27052004_162842_2168/app_query.fasta_1.647
-DB=Issued Patents NA -QMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTEM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0902772@cgn 1.1.69 @runat27052004_162842_2168 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCFUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2444	100.0	1447	3	US-08-878-177-1		Sequence 1, Appli
2	2414.5	98.8	1528	3	US-08-878-177-3		Sequence 3, Appli
3	1624.5	66.5	2938	2	US-08-343-443B-3		Sequence 3, Appli
4	511	20.9	1752	3	US-09-360-779-1		Sequence 1, Appli
5	511	20.9	1752	4	US-09-435-335-1		Sequence 1, Appli
6	449	18.4	1604	1	US-08-306-691B-43		Sequence 43, Appli
7	449	18.4	1604	5	PCT-US93-06251-9		Sequence 9, Appli
8	445.5	18.2	2268	3	US-09-344-579-1		Sequence 1, Appli
9	407	16.7	2532	4	US-09-620-312D-869		Sequence 869, App
10	342.5	14.0	2667	2	US-08-469-412A-1		Sequence 1, Appli
11	342.5	14.0	2667	3	US-09-021-715-1		Sequence 1, Appli
12	322	13.2	2064	3	US-08-875-944B-1		Sequence 1, Appli

13	322	13.2	2064	3	US-09-116-049-3		Sequence 3, Appli
14	322	13.2	2064	4	US-09-602-868A-1		Sequence 1, Appli
15	322	13.2	2064	4	US-09-884-363-3		Sequence 3, Appli
16	306	12.5	2410	2	US-08-780-835B-1		Sequence 1, Appli
17	306	12.5	2410	3	US-08-303-368-1		Sequence 1, Appli
18	306	12.5	2410	3	US-09-116-049-1		Sequence 1, Appli
19	306	12.5	2410	4	US-09-884-363-1		Sequence 1, Appli
20	306	12.5	2410	4	US-09-328-525-13		Sequence 13, Appli
21	301.5	12.3	328	2	US-08-343-443B-5		Sequence 5, Appli
22	301	12.3	1894	4	US-09-570-593-1		Sequence 1, Appli
23	301	12.3	1905	3	US-09-055-113-2		Sequence 2, Appli
24	301	12.3	3317	4	US-09-570-593-12		Sequence 12, Appli
25	291.5	11.9	2266	2	US-08-213-767-1		Sequence 1, Appli
26	279.5	11.4	1933	4	US-09-920-759-3		Sequence 3, Appli
27	279.5	11.4	1976	4	US-09-920-759-10		Sequence 10, Appli
28	273.5	11.2	5510	3	US-09-009-913-3		Sequence 3, Appli
29	261.5	10.7	5427	3	US-09-009-913-2		Sequence 2, Appli
30	261.5	10.7	5667	3	US-09-009-913-4		Sequence 4, Appli
31	259	10.6	665	4	US-09-920-759-11		Sequence 11, Appli
32	234.5	9.6	2544	2	US-08-469-412A-6		Sequence 6, Appli
33	234.5	9.6	2544	3	US-09-021-715-6		Sequence 6, Appli
34	233	9.5	731	4	US-09-976-594-191		Sequence 191, App
35	231.5	9.5	848	3	US-09-009-913-338		Sequence 338, App
36	231.5	9.5	2280	3	US-09-009-913-8		Sequence 8, Appli
37	231.5	9.5	2428	3	US-09-009-913-6		Sequence 6, Appli
38	231.5	9.5	2498	3	US-09-009-913-10		Sequence 10, Appli
39	230	9.4	1907	4	US-09-300-958A-27		Sequence 27, Appli
40	230	9.4	1907	4	US-09-570-593-4		Sequence 4, Appli
C 41	229	9.4	852	3	US-09-020-856-44		Sequence 44, Appli
C 42	229	9.4	852	3	US-09-030-607-44		Sequence 44, Appli
C 43	229	9.4	852	4	US-09-439-313-44		Sequence 44, Appli
C 44	229	9.4	852	4	US-09-352-616A-44		Sequence 44, Appli
C 45	229	9.4	852	4	US-09-232-149A-44		Sequence 44, Appli

ALIGNMENTS

RESULT 1
US-08-878-177-1
; Sequence 1, Application US/08878177

; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et

; GENERAL INFORMATION:

; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of

; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001

; CURRENT APPLICATION NUMBER: US/08/878,177

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1447

; TYPE: DNA

; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA

US-08-878-177-1

Alignment Scores:

Pred. No.: 2.54e-249 Length: 1447
Score: 2444.00 Matches: 451
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-902-772-2 (1-451) x US-08-878-177-1 (1-1447)

QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20

63 ATGGCAGACACTATTAAAGGAGCATTTATCAGTGGTGTGAGTGAAGACAGCTCTTTTGG 122

QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40

123 TGTGCTACGGATGGCCCACTTGCAAGACAGAAATGACGCCCTCTCTCCATGTA 182

Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
Db 183 TATGGGCAACATCAAGATGAGCCCGCGTTCGCCAGCAGACTGGTATATCAAGGCC 242

Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 243 CCGGCCAGAGTTACCATTAAGATGGAGTGTACCCCAACACAGGTTAATGGGTCAAGGAAT 302

Qy 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db 303 TCACCTGATGACGTGCGCGTGGCAAGAGGAGGAAATGGTTAGCAGTTTCAGACAAATGTT 362

Qy 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120
Db 363 GGGATGAACTATGGAGCTACATGGAGAGAGCATATTCGGCTCCAAATATGACACC 422

Qy 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
Db 423 AATCAACGAAGAGTTATTTGTCGCCAGAGATCTTACCGTTATGGAGCAGACACCATGTACGG 482

Qy 141 GlnTyrLeuGlnThrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
Db 483 CAGTGGCTGGAGTGGGAGTGAAGAGTATGGTCTCCAGACGTGACATCTTGTGTTC 542

Qy 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db 543 CAGAACATTCATGGGAAGAGTGTCTGTAATGACCAAGATGACATTCACAGAGACTCACG 602

Qy 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAla 200
Db 603 CCGAGCTATTAACGAGATATCTCCCTGTCACACTACCTACCTCAGAGAGAGAGGCC 662

Qy 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220
Db 663 ACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAAGCAAGAAATACAAACAAG 722

Qy 221 ProAspLeuProTyrGluGlnAlaArgArgSerAlaTyrThrSerHisSerHisProThr 240
Db 723 CCAGATTACCTTATGAGCAAGAGAGAGATCAGCGTGGAGCTCAGAGCCATCCCAT 782

Qy 241 GlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAspGlnArgPro 260
Db 783 CAGTCAAAAGCTACCCNACCATCATCTTCAACAGTGCCTCCCAACAGACACAGCGTCT 842

Qy 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
Db 843 CAGTTAGATCCTTATCAGATCTTGGACCGCAGCAGCGCTCTTGCAAAATCCAGGGAGT 902

Qy 281 GlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsn 300
Db 903 GGGCAGATACAGTATGGCGAGTTCCTACTGGAGCTTCTGCGACAGCTCCAACTCCAC 962

Qy 301 CysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320
Db 963 TGCATACCTGGGAGGCAACAAATGGGAGTTCAGATGACAGACCTGATGAAGTGGCT 1022

Qy 321 ArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
Db 1023 CGGCGTTGGGAGAGAGAAAGAAACCTTAACATGAACTATGACAACTCAGCCGTGCA 1082

Qy 341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
Db 1083 CTTGCTGCTACTATGACAAAAATATTATGACTAAAGTTTCATGGTAAACGCTATGCTAC 1142

Qy 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerSerMet 380
Db 1143 AAATTTGATTTCCACGAATCGCTCAGCGCCCTCAGCCCTCAGCCCTCAGAAATCATCATG 1202

Qy 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
Db 1203 TACAAATACCATCAGACCTCCCTCATATGATGTTCTTCCATGACACACCCCAAGAGATG 1262

Qy 401 AsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhePheAlaAla 420
Db 1263 AACTTTTGTAGTCCCATCCCTGCTTTGCCGTAACTCACTCACTTTTGTGCTGCC 1322

Qy 421 ProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAla 440
Db 1323 CCTAATCCATCTGGAATTCACCACTGGAGGATCTACCCCAATACCAAGCTGCCAGCT 1382

Qy 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1383 GCTCATATGCTTCCCATCTTGGCACCTACTAC 1415

RESULT 2
US-08-878-177-3
; Sequence 3, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878.177
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: DNA
; ORGANISM: c-erg gene, chicken DNA
US-08-878-177-3

Alignment Scores:
Pred. No.: 3,74e-246 Length: 1528
Score: 2414.50 Matches: 450
Percent Similarity: 94.14% Conservative: 0
Best Local Similarity: 94.14% Mismatches: 1
Query Match: 98.79% Indels: 27
DB: 3 Gaps: 1

US-09-902-772-2 (1-451) x US-08-878-177-3 (1-1528)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 63 ATGCGACGACATTAAGGAAGCATTTATCAGTGTGTAGTGAAGACCATCTCTTTGTCAG 122

Qy 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
Db 123 TGTGCTACGGATCGCCCCACCTTGCAGAGACAGAAATGACAGCTCTCTTCCAGTGA 182

Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
Db 183 TATGGCAACATCAAGATGAGCCCGCTTCCCGCAGGAGCTGGTTATCACAGCCC 242

Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 243 CCGGCCAGAGTTACCATTAAGATGGAGTGTAAACCCAAACCAAGGTTAATGGGTCAAGGAAT 302

Qy 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db 303 TCACCTGATGATCGACGCTGGCAAGAGGAGGAAATGGTTAGCAGTTTCAGACAAATGTT 362

Qy 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120
Db 363 GGGATGAACTATGGAGCTACATGGAGAGAGCATATTCGGCTCCAAATATGACAACT 422

Qy 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
Db 423 AATGAACGAAGAGTTATTTGTCGCCAGAGATCTTACGTTATGGAGCAGACACCATGTACGG 482

Qy 141 GlnTyrLeuGlnThrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
Db 483 CAGTGGCTGGAGTGGGAGTGAAGAGTATGGTCTTCCAGACGTGACATCTTGTGTTC 542

```

QY 161 GlnAsnIleaspGlyLysGluLeuCysLysMetThrLysaspPheGlnArgLeuThr 180
Db 543 CAGAACATTGATGGAAAGATTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGC 602
QY 181 ProSerTyvAsnAlaAspIleLeuLeuSerHisLeuHisTyvLeuArgGluArg 198
Db 603 CCGAGCTATAACGAGATATCTCTGTGCACACCTACCTACCTCAGAGAGACTCCTCTT 662
QY 198 ----- 198
Db 663 CCACATTGACTTCAGATGATGTTGATAAGCCCTTACAAACTCTCCACGGTTAATGCAT 722
QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyvProGluAla 213
Db 723 GCTAGAACACAGAGAGGCCCTTTATTTTTCRAATACATCAGTTTACCCAGAGCA 782
QY 214 ThrGlnArgIleThrArgProAspLeuProTyvGluGlnAlaArgAgsSerAlaTyr 233
Db 783 ACGCAAAAGATAAACAAGCCAGATTACCTTATGAGCAAGCGAGAGATCAGCGTGG 842
QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 253
Db 843 ACGAGTCACGCCATCCCATCTCACTCAAGCTACCCACCATCATCTTCAACAGTCCCC 902
QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyvGlnIleLeuGlyProThrSerSer 273
Db 903 AAAACAGAGACCCAGCGTCTCAGTTAGATCTTATCAGATCTTGGACCGACGAGCAGC 962
QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuGluLeuLeu 293
Db 963 CGTCTTGAAATCCAGGAGGAGGCGAGATACAGTATGGCAGTTCTTACTGGAGCTTCG 1022
QY 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
Db 1023 TCGACAGCTCCACTCCACTGCATGCATCCCTGGAGGGCAAAATGGGAGTTCAAGATG 1082
QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333
Db 1083 ACAGACCTGTATGAATGGCTCGCGTGGGAGAGAGAGAAAGCAAAACCTAACATGAAC 1142
QY 334 TyrAspLysLeuSerArgAlaLeuArgTyvTyvTyvAspLysAsnIleMetThrLysVal 353
Db 1143 TATGCAAACTCAGCGGTGCTGCTGCTACTACTATGCAAAAATATATGACTAAAGTT 1202
QY 354 HisGlyLysArgTyvAlaTyvLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
Db 1203 CATGTAACGCTATGCTCAAAATTTGATTTCCAGCGAATCGCTCAGGCGCTCCAGCT 1262
QY 374 HisProGluSerSerSerMetTyvLysTyvProSerAspLeuProTyvMetSerSerTyv 393
Db 1263 CACCTCCAGAAATCATCATGTAATAATCCCATCAGACCTCCCTCATGAGTTCTCTAC 1322
QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413
Db 1323 CATGCACACCCCCAGAGACATGAATCTTGTAGCTCCCATCCCTGCTTGTCCCGTAACC 1382
QY 414 SerSerSerPhePheAlaAlaProAsnProTyvTrpAsnSerProThrGlyGlyLysTyv 433
Db 1383 TCATCCAGCTTTTGTGTCGCCCTTAATCCATCTAGTGAATCCCACTGGAGGATCTAC 1442
QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyvTyv 451
Db 1443 CCCAATACAGGCTGCCAGCTGCTCATATGCTTCCCATCTTGGCACCTACTAC 1496

```

RESULT 3

```

US-08-343-443B-3
; Sequence 3, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Melot, Thomas

```

```

; APPLICANT: Peter, Martine
; APPLICANT: Ploougastel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989,6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2938 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 143..1498
; US-08-343-443B-3

```

```

Alignment Scores:
Pred. No.: 6,89e-162 Length: 2938
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
DB: 2 Gaps: 7
US-09-902-772-2 (1-451) x US-08-343-443B-3 (1-2938)

```

```

QY 1 MetAlaSerThrLysLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 143 ATGACCGGACATTAAAGGAGGCTCTGTCGGTGTGAGCGACGACGACGACGACGAC 202
QY 21 CysAlaTyvGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 203 TCAGCGTAGGAGCGGCGGACGCCCATCTCCCCAGCGCGGACATGACTGCTCCGCGGAGTCT 262
QY 40 GluTyvGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
Db 263 GACTACGGGAGCGCCGCCACAGATCAACCCCTCCACACGACGAGGAGTGGATCAATCAG 322
QY 60 ProProAlaArgValThrLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79

```

Db 323 ---CGAGTGGGTCACAGTCAAGCGGAGTAT-----GACCACATGAATGATCCAGG 373
QY 80 AsnSerProAspAspCysSerValAlaIysGlyLysMetValSerSerSerAspAsn 99
Db 374 GAGTCTCCGGTGGAGTCCAGCGTGTAGCAATGACAGAGCTGTGGCGGAGCGAGTCC 433
QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis--1leProProProAsnMet 118
Db 434 AACCCCATGAACACACAGCTATATGACGAGAGAGTGGCCCCCTCTCTCCACATG 493
QY 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 494 ACCACCAACGAGAGAGAGTATCGTCCCGCAGACCCACACTGTGTACACAGAGAGCAT 553
QY 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
Db 554 GTGAGCAATGGCTGGAGTGGCCCATAAAGAGATATAGCTTGATGGAGATCGACATCC 613
QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 614 TTTTTCAGAACATGATGGACAGAACTGTGTAAATGAACAGAGAGAGTCTCTCCCGC 673
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
Db 674 GCCACACCCCTCTACACACGGAAGTGTGTGTACACCTCAGTTACCTCAGGAA--- 730
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db 731 --AGTTCACCTGCTGCTATATAACACCTCCACACAGCAATCTCTCAGGATGAGT 787
QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHis 238
Db 788 GTCAAAAGAACCCCTCTTATGATCATGTCAGAGAGAGAGTGGGCAATACATGAAT 847
QY 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
Db 848 TCTGGCCTCAACAAAGTCTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 907
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 908 CAACGGCCCCAGCAGATCCGATACAGATCTGGGCCCGCAGCAGTCCGCTTAGCCAAC 967
QY 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeuSerAspSerSer 297
Db 968 CTTGGAAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGAGGCTCTCTCGACAGCGCC 1027
QY 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 1028 AACGCCAGCTATACCTGAGGGGACCAACAGGGGAGTTCAAAATGACGGACCCCGAT 1087
QY 318 GluValAlaArgAspTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 1088 GAGTGGCCAGCGCTGGGGCGAGCGGAAAGCAGCCCAACATGAATACGACAGCTG 1147
QY 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1148 AGCGGGGCCCTCCGTTATTATATGATAAAACATTTATGACCAAAAGTGCACGGCAGAA 1207
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu 377
Db 1208 TATGCTTACAAATTTGATCTCCAGCGCATTCAGGCTCTGACGACATCCGACCGAG 1267
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1268 TCGTCCATGTACAAAGTACCTCTGACATCTCTTACATGCTCTCTACCATGCCACAG 1327
QY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
Db 1328 CAGAAGGTGAATTTGTCTCCCTCCCATCTCCATCTCCAGTCTGTCATCTCTCCAGCTTC 1387
QY 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435

Db 1388 TTTGAGCGGCATCACAATACTGGACCTCCCCACGGGGGAATCTACCCCAACCCCAAC 1447
QY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1448 GTCCCCCGGCATCTTAACACCCACGTCCTTACACTTAGGAGCTACTAC 1498
RESULT 4
US-09-360-779-1
; Sequence 1, Application US/09360779
; Patent No., 6266216
; GENERAL INFORMATION:
; APPLICANT: Deneris, Evan S. V.
; APPLICANT: Fyodorov, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
; FILE REFERENCE: CASE-03828
; CURRENT APPLICATION NUMBER: US/09/360,779
; CURRENT FILING DATE: 1999-07-26
; EARLIER APPLICATION NUMBER: 60/094,264
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(1131)
US-09-360-779-1
Alignment Scores:
Pred. No.: 3,849-44 Length: 1752
Score: 511.00 Matches: 123
Percent Similarity: 50.69% Conservative: 23
Best Local Similarity: 42.71% Mismatches: 63
Query Match: 20.91% Indels: 79
Gaps: 10
US-09-902-772-2 (1-451) x US-09-360-779-1 (1-1752)
QY 209 ValTyrProGluAlaThrGlnArgIleThrArgProAspLeuProTyrGluGlnAla 228
Db 369 GTGCCCGCCCGGCGGTCCTCCCATCTCCACCCCA----- 407
QY 229 ArgArgSerAlaTrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSer 248
Db 408 -----GTCCCCGGCAGCATGAGCAGAGCGGCACCTCCAGCCCTGCT 452
QY 249 SerSerThrValProLysThrGluAspGlnArg-----Pro 260
Db 453 GATCAACATGTACCTACCATCCGTCGAGATGCTCTTTTAAAGGAAGGAGAGGCC 512
QY 261 GlnLeu-AspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySe 280
Db 513 GAGCTGGGGGCGG-----CTGAGCCCTCGGTACAGAAA-----GGCAG 551
QY 280 rGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeuSerAspSerSerAsnSerAs 300
Db 552 CGGGCAGATCCAGTTGTGGCAGTTTCTACTGAGCTGTGGCAGACCGCGCAGCGCG 611
QY 300 nCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluVal 320
Db 612 CTGCATCGCTGGAGGGCGGCGCAGCGGAGTTCAGCTCACCAGCCCGCAGCGGTGGC 671
QY 320 aArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAl 340
Db 672 GCGAGCTCGGCGGAGCGCAGAGCAAGCCCAATATGACTACGACAGCTAAGTCGAGC 731
QY 340 aLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTy 360
Db 732 ACTGGGCTACTACTACGACAAACATCATGAGCAAGGTGCGCAGCGCTACGCTTA 791


```
QY 360 rLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisPro----- 375
Db 792 CCGCTTTGACTTCAGGCGCTGGACAGGCTTGCAGCCACCACCGCCGCGCAGCCCGC 851
QY 376 -----ProGluSerSerMetTyrLysTyrPr 384
Db 852 CGCGCTGCCGCGCGCAGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 911
QY 384 oSerAspLeu-----ProTyrMetSerSerTyrHisAlaHisProGlnTy 399
Db 912 GGCTGGCTGGCTCCACTGCCCTTCCCGCGCGCTCC-----AA 950
QY 399 sMetAsnPheValAlaProHisProAlaLeuLeuProValThrSerSerSerPheAl 419
Db 951 ACTCAACCTTATGCGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 999
QY 419 aAlaProAsnProTyrTrpAsnSerPro-----ThrGlyGl 431
Db 1000 -----TACTGGCTGTGCCCAAGCGCAGCGCGCGCTGCCCGCGCGCGCG 1046
QY 431 yIleTyrProAsnThrArgLeuPro-----AlaAlaHisMetPr 444
Db 1047 GCTTACCACCGCGCGCGCTTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1106
QY 444 oSerHisLeuGlyThrTyrTyr 451
Db 1107 TTCGCACTTGGGGGTGCATTAT 1128

RESULT 5
; US-09-435-335-1
; Sequence 1, Application US/09435335
; Patent No. 6384204
; GENERAL INFORMATION:
; APPLICANT: Deneris, Evan S.
; APPLICANT: Pyodoro, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
; FILE REFERENCE: CASE-04027
; CURRENT APPLICATION NUMBER: US/09/435,335
; CURRENT FILING DATE: 1999-11-05
; EARLIER APPLICATION NUMBER: 09/360,779
; EARLIER FILING DATE: 1999-07-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(1131)
; US-09-435-335-1

Alignment Scores:
Pred No.: 3,84e-44 Length: 1752
Score: 511.00 Matches: 123
Percent Similarity: 50.6% Conservative: 23
Best Local Similarity: 42.71% Mismatches: 63
Query Match: 20.91% Indels: 79
DB: 4 Gaps: 10

US-09-902-772-2 (1-451) x US-09-435-335-1 (1-1752)
QY 209 ValTyrProGluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAla 228
Db 369 GTGCCCGCGCGTGGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 407
QY 229 ArgArgSerAlaThrThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSer 248
Db 408 -----GTCCCGCGCAGCGCATGACAGACAGCGCGCGCGCGCGCGCGCGCGCG 452
```

```
QY 249 SerSerThrValProLysThrGluAspGlnArg-----Pro 260
Db 453 GATCAACATGTTACTTACAGATCCCGTGGAGATGGTCTTTTAAAGGAAGGAGAGCC 512
QY 261 GlnLeu-AspProTyrGlnIleLeuLeuGlyProThrSerSerArgLeuAlaAsnProGlySe 280
Db 513 GAGCTGGGGCGCG-----CTGAGCGCTCGGTACAGAAA-----GGCAG 551
QY 280 rGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 300
Db 552 CGGCGAGATCCAGTTGTGGCAGTTTCTACTGGAGTGTCTGGCAGCGCGCGCGCGCGCG 611
QY 300 nCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAl 320
Db 612 CTGCATCGCTGGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 671
QY 320 aArgArgTGTGlyGlyArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAl 340
Db 672 GCGAGCTGGGGCGAGCGCGAGAGCAAGCCCATATGAACTACGACAGCTAAGTCGAGC 731
QY 340 aleuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTy 360
Db 732 ACTCGCTACTACTACGACAAAAACATCATGAGCAAGGTGCAAGCGCAGCGCTACGCTA 791
QY 360 rLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisPro----- 375
Db 792 CCGCTTTGACTTCCAGCGCGCTGCCAGCGCTTGCAGCGCGCGCGCGCGCGCGCGCG 851
QY 376 -----ProGluSerSerMetTyrLysTyrPr 384
Db 852 CGCGCTGCCGCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 911
QY 384 oSerAspLeu-----ProTyrMetSerSerTyrHisAlaHisProGlnTy 399
Db 912 GGCTGGCTGTGCCACTGCCCTTCCCGCGCGCTCC-----AA 950
QY 399 sMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhePheAl 419
Db 951 ACTCAACCTTATGCGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 999
QY 419 aAlaProAsnProTyrTrpAsnSerPro-----ThrGlyGl 431
Db 1000 -----TACTGGCTGTGCCCAAGCGCAGCGCGCGCGCGCGCGCGCGCGCGCG 1046
QY 431 yIleTyrProAsnThrArgLeuPro-----AlaAlaHisMetPr 444
Db 1047 GCTTACCACCGCGCGCGCTTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1106
QY 444 oSerHisLeuGlyThrTyrTyr 451
Db 1107 TTCGCACTTGGGGGTGCATTAT 1128

RESULT 6
; US-08-306-691B-43
; Sequence 43, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
```

```

; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: NO. 5734039e
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1604 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-306-691B-43

Alignment Scores:
Pred. No.: 1,25e-37 Length: 1604
Score: 449.00 Matches: 114
Percent Similarity: 43.88% Conservative: 51
Best Local Similarity: 30.32% Mismatches: 89
Query Match: 18.37% Indels: 122
DB: 1 Gaps: 9

US-09-902-772-2 (1-451) x US-08-306-691B-43 (1-1604)

Qy 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 447 ACTAAGAACAGCAACGACTGGGATCCCAAGAGACCCCGGAGTGGACAGAAACCAT 506
Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTrpGlyLeuProAspValIleLeu 158
Db 507 GTTCGGGAGCTGGGAGTGGGCTGTGAATATCAGCTGAAGAGTGTAGACTTCCAG 566
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 567 AAGTTC---TGATGATGAGCAGACCCCTCTGGCCCTGGTAAAGACTGCTTTTCGAG 623
Qy 179 LeuThrProSerThrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlu--- 197
Db 624 CTGCCCCAGACTTTGTTGGGACATCTTATGGGAACATCTAGAGATCCTGCAAGAGAG 683
Qy 198 -----ATGGlyAlaThrPheIlePheProAsnThr----- 207
Db 684 GATGTGAACCATATCAAGTTAATGGAGTCAACCCAGCTATCCAGATCCCGCTATACC 743
Qy 207 ----- 207
Db 744 TCGGATTACTTCAATAGCTATGATTGAGCATGCCAGTGTGTTCACCATCCGAGTTC 803
Qy 208 -----SerValTyrProGluAlaThrGlnArg 216
Db 804 TCAGAGCCAGCTTCATCAGAGTCTCTATCAGACGCTCCATCCCATCAGCTCGGAGAG 863
Qy 217 Ile----- 217
Db 864 CTCCTCTCCCTCAAGTATGAGATGACTACCCCTCGGTCTATTTCTCCGAGACCCCTCTCCAG 923
Qy 217 ----- 217
Db 924 ACAGACACCTTGCAGATGACTACTTTGCTATCAACAAGAGTCTGCACCCAGACAC 983
Qy 218 -----ThrThrArgProAspLeuProTyrGluGlnAlaArgSer--- 231
Db 984 ATGTGCATGGGGAGGACCGAGTCGTGGTAAACTCGGGGGCCAGGACTCTTTTGAAGCATA 1043

```

```

Qy 232 ---AlaTrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSer 250
Db 1044 GAGAGCTACGATAGTGTGATCGCTCACCAGTCTCTGGAGCAGCCAGTCACTTTCAAC 1103
Qy 251 ThrValProLys-----ThrGluAsp----- 257
Db 1104 AGCCTGACGGGTGTTCCTCTATGACAGCTTCGACTCAGAGGACTATCGGCTGCCTG 1163
Qy 258 -----GlnArgProGlnLeuAsp 263
Db 1164 CCCAACCAAGCCCAAGGGCACCTTCAAGGACTATGTGGGGACCGTGTGCTACCTCAAT 1223
Qy 264 ProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaIleProGlySerGlyGlnIle 283
Db 1224 AAGGACAAAGCCTGCTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1283
Qy 284 GlnLeuTrpGlnPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 303
Db 1284 CAGCTATGGCAGTTTCTTCTGGAATTAATCTACTGATTAATCTCTAGTCTCTCTTATCAGC 1343
Qy 304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArgTrp 323
Db 1344 TGGACAGGAGATGGCTGGGAATTCAAACTTCTGACCCAGATGAGTGGCCAGGAGATGG 1403
Qy 324 GlyGluArgLysSerLysProAsnMetAsnTyrAspLysLysLeuSerArgAlaLeuArgTyr 343
Db 1404 GGAAGAGGAAAAACAAACCTTAAGATGAATATGAGAACTTGAGCCGTGGCTACGCTAC 1463
Qy 344 TyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe--- 362
Db 1464 TATTACGACAAACATCATCCACAGACAGCGGGGAAACGCTACGTGTACCGCTTTGTG 1523
Qy 363 ---AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1524 TGTGACCTG-----CAGAGCTGTGGGTACACCCCTGAG 1559

RESULT 7
PCT-US93-06251-9
; Sequence 2, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1604 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-9

Alignment Scores:
Pred. No.: 1,25e-37 Length: 1604
Score: 449.00 Matches: 114
Percent Similarity: 43.88% Conservative: 51
Best Local Similarity: 30.32% Mismatches: 89
Query Match: 18.37% Indels: 122
DB: 5 Gaps: 9

US-09-902-772-2 (1-451) x PCT-US93-06251-9 (1-1604)
QY 119 ThrAsnGlnArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
DB 447 ACTAAGACAGACAGACGCTGGGATCCCAAGAGACCCCGGAGTGGACAGAACCCAT 506
QY 139 ValArgGlnTrpLeuGlnTrpAlaValLysGluTrpGlyLeuProAspValAspLeu 158
DB 507 GTTCGGGACTGGGTGATGTGGCTGTGAATGAATTCAGCTGGAAGAGGTGTAGACTTCCAG 566
QY 159 LeuPheGlnAsnLeuAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
DB 567 AAGTTC---TGTATGAATGGAGCAGCCCTCTGGCCCTGGGTAAAGACTGCTTCTCGAG 623
QY 179 LeuThrProSerTyAsnAlaAspIleLeuLeuSerHisLeuHisTyLeuArgGlu--- 197
DB 624 CTGGCCCCAGACTTGTGTGGGACATCTTATGGAAACATCTAGAGATCTCTGCAGAAAGAG 683
QY 198 -----ArgGlyAlaThrPheIlePheProAsnThr----- 207
DB 684 GATGTGAACCATATCAAGTTAATGGAGTCAACCCAGCCTATCCAGATCCCGCTATACC 743
QY 207 ----- 207
DB 744 TCGGATTACTTTCATAGCTATGTTGATGAGTATGCCAGTCCCGGCTGTTCCACCATCGGAGTTC 803
QY 208 -----SerValTyProGluAlaThrGlnArg 216
DB 804 TCAGAGCCAGCTTTCATCAGAGATCCTATCAGAGCTCCATCCCATCAGCTCGGAAGAG 863
QY 217 Ile----- 217
DB 864 CTCCTCTCCCTCAAGTATGAGATGACTACCCCTCGGTCAATCTCCGAGACCTCTCCAG 923
QY 217 ----- 217
DB 924 ACAGACACCTTCAGAGATGACTACTTGTGCTATCAACAGAGTTCGTCACCCAGACAAC 983
QY 218 -----ThrThrArgProAspLeuProTyArgGluGlnAlaArgAspSer--- 231
DB 984 ATGTGATGGGAGGACAGTCGTTGGTAACTCGGGGGCCAGGACTCTTTTGAAGCATA 1043
QY 232 ---AlaTrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSer 250
DB 1044 GAGAGCTACGATAGTGTGTGCTGCTCACCAGTCTCCGAGCAGCCAGTCACTCTTCAAC 1103
QY 251 ThrValProLys-----ThrGluAsp----- 257
DB 1104 AGCCTGACGGGTGTCCTCCTATGACAGCTTCGACTCAGAGGACTATCGGCTGGCCCTG 1163
QY 258 -----GlnArgProGlnLeuAsp 263
DB 1164 CCCAACACAGCCCAAGGACCTTCAGGACTATGTGGGACCTGTGACCTCAAT 1223
QY 264 ProTyGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIle 283
DB 1224 AAGGACAGCCTGCTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1283
QY 284 GlnLeuTrpGlnPheLeuLeuLeuLeuSerSerSerSerSerSerSerSerSerSer 303
DB 1284 CAGTATGGCAGTCTTCTTCTGGAATTACTACTGATATAAATCTCTGCTGCTGCTGCTGCT 1343
QY 304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgTrp 323
DB 1344 TGGACAGAGATGCTGGGAATTCAACTTCTGACCCAGATGAGTGGCCAGGAGATGG 1403
QY 324 GlyGluArgLysSerLysProAsnMetAsnTyAspLysLeuSerArgAlaLeuArgTy 343
DB 1404 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1463
QY 344 TyTrpAspLysAsnIleMetThrLysValHisGlyLysAspTyAlaTyLysPhe--- 362
DB 1464 TATTACGCAAAACATCATCTCCACAGACCGGGGAAACGCTACGCTACCGCTTGTG 1523
QY 363 ---AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
DB 1524 TGTGACCTG-----CAGAGCCTGCTGGGTACACCCCTGAG 1559

RESULT 8
US-09-344-579-1
; Sequence 1, Application US/09344579
; Patent No. 6054316
; GENERAL INFORMATION:
; APPLICANT: Brenda F. Baker
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF ETS-2 EXPRESSION
; FILE REFERENCE: R1S-0063
; CURRENT APPLICATION NUMBER: US/09/344,579
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (291)..(1700)
US-09-344-579-1

Alignment Scores:
Pred. No.: 5.2e-37 Length: 2268
Score: 445.50 Matches: 127
Percent Similarity: 40.96% Conservative: 43
Best Local Similarity: 30.60% Mismatches: 105
Query Match: 18.23% Indels: 141
DB: 3 Gaps: 14

US-09-902-772-2 (1-451) x US-09-344-579-1 (1-2268)
QY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
DB 570 CAGCGCGCTGGGCATTCCTCAAGAACCCCTGGCTGGGAGTGAACAACAGGTATGCCAG 629
QY 142 TrpLeuGluTrpAlaValLysGluTyArgGlyLeuProAspValAspIleLeuPheGln 161
DB 630 TGGCTTCTTGGGCCACCAATGATGTTGCTGGTGAACCTGCTGAGAGTTC--- 686
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB 687 GCATGAATGGCCAGATGCTGTGTAACCTTGGCAAGGAAACGCTTCTGGAGCTGCACCT 746
QY 182 SerTyAsnAlaAspIleLeuLeuSerHisLeuHisTyLeuArgGluArgGlyAlaThr 201
DB 747 GACTTGTGGTGACATTCCTGGGAACATCTGGGACCAATGATCAAGAAACCA--- 803
QY 202 PheIlePheProAsnThrSerValTyProGluAlaThrGlnArgIleThrThrArgPro 221
DB 804 -----GAAAGACAGAGATCAATATGAAGAAATTCACACTCACCTCCCTTCT 854
QY 222 -----AspLeuProTyArgGlnAla 228
DB 855 CATGTGATTACAGCAATACATTAGTTTGTGGCAGACGCGCCCTATGGAATGCGAG 914

```

```

QY 228 aAIGaG-----SerAlaThrThrSer-----HisSerHisProTh 240
Db 915 ACACAGAAATACCCAAAGCGCGCTCTGGHACAGCATGTGTCCGCGCTCCACACCCAGC 974
QY 240 rLnSerIysAlaThrGln-----ProSerSe 249
Db 975 GTACTCAGCTCTGAGCAGGAGTTTCAGATGTTCCCAAGTCTCGGCTCAGCTCGTCCAGC 1034
QY 249 rSerThr----- 251
Db 1035 GTCACTACTGCTCTGCTCAGTCAGCACTTCCAGCAGCACTTGAATTGCTCACCAC 1094
QY 252 -----ValProLysThr-----GluAs 257
Db 1095 AATCTGGGACTCCCAAGACACAGCACTCCCTGAGAACGGTGGGACAGCTTCGAGAGC 1154
QY 257 pGlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270
Db 1155 TCAGACTCCCTCCTCCAGCTCTGGGAACAGCAGCTGCTGCTGGATGGCAACGGTT 1214
QY 270 ----- 270
Db 1215 CTTCTCTGAGAGCTTCCGAAGATGACTGCGACCCAGCTCTCTGCTCCTCAATAAGCCAAC 1274
QY 271 -----ThrSerSerArgLeuAlaAsnPro----- 278
Db 1275 ATGCTTTCAAGGATATCATCAAGAGAGG-AGTGACCCAGCTGGAGCAAGCAACAGT 1333
QY 279 -----GlySerGlyGlnIleGlnLeuThrGlnPh 288
Db 1334 TATACCTGAGCTGTGCTGGCGGCTTCCACAGGAAGTGGACCTATTGAGCTGTGGCAGTT 1393
QY 288 eLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAs 308
Db 1394 TCTCTGAGAGCTGCTATCAGCAAAATCTGCAGCTATTCTACAGCTGGAGCTGAGAGCGG 1453
QY 308 nGlyLeuPheIysMetThrAspProAspGluValAlaArgTrpGlyGluArgLysSe 328
Db 1454 ATGGGAGTTTAACTCTGCGCGACCCCGATGAGTGGCGCGCGCGGGGAAAGGAGAAA 1513
QY 328 rIysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAs 348
Db 1514 TAAAGCCAGATGAATCAGAGAGCTGAGCGGGGCTTACGCTACTATTACGACAGAA 1573
QY 348 nIleMetThrIysValHisGlyLysArgTyrAlaTyrLysPhe----- 362
Db 1574 CATCATCCACAGAGCTCGGGGAAGCGCTACGTGTACCGTTCGTGTGCGACCTCCAGAA 1633
QY 363 -----AspPheHisGlyIleAlaGlnAlaLeuGlnProHisPr 375
Db 1634 CTTGCTGGGTTCACCGCCGAGGAAGTCAAGCCATCTG---GGCGTCCAGCGCCGACAC 1690
QY 375 oProGluSerSerMetTyrLysTyrProSerAspLeuPro---TyrMetSerSerTyrHi 394
Db 1691 GGAGGACTGAGTCTGCGCGGACCACTGAGCGCGCGCGAGGCTCGTGGACTGAGTGGGA 1750
QY 394 sAlaHisProGlnLysMetAsnPheValAlaProHisProPro 408
Db 1751 AGCCCATCTCGACAGCTG-----CTCCCG 1775

```

RESULT 9

```

US-09-620-312D-869
; Sequence 869, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom

```

```

; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pc_FL_genes Version 1.0
; SEQ ID NO 869
; LENGTH: 2532
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2532)
; US-09-620-312D-869

Alignment Scores:
Pred. No.: 7,56e-33 Length: 2532
Score: 407.00 Matches: 94
Percent Similarity: 56.67% Conservative: 16
Best Local Similarity: 56.97% Mismatches: 23
Query Match: 16.65% Indels: 34
DB: 4 Gaps: 5

US-09-902-772-2 (1-451) x US-09-620-312D-869 (1-2532)

QY 271 ThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeu 290
Db 2073 ACCCTCGAAGAGAAAGAAATGATGACGGGACACATCAGCTGGGAGGATTTCTGCTG 2132
QY 291 GluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGlu 310
Db 2133 GAGCTGCTGCTGACCGCGGAACCGGCG-TGCATCGCTGGAGGGCGGTCAAGCGCAG 2191
QY 311 PheLysMetThrAspProAspGluValAlaArgTrpGlyGluArgLysSerLysPro 330
Db 2192 TTCAGCTCACGACCGGACGAGGTGGCGCGCGGTGGGCGAGCGCAAGAGCAAGCCC 2251
QY 331 AsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMet 350
Db 2252 AACATGAATACGACAAAGCTGAGCGCGCGCT-ATCTACTACGACAAAGAACATCATG 2310
QY 351 ThrIysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAla 370
Db 2311 AGCAGGTGATGTGTAAGCTACCGCTTACGCTTTCAGGCTTCAGGCTTCGCGAG--- 2367
QY 371 LeuGlnProHisProGluSerSerMetTyrLysTyrProSerAspLeu----- 387
Db 2368 -----GACGCGCGCGCTCTCAAGCTGCCGCGCGCGCTCCGACCGCTG 2409
QY 388 -----ProTyrMetSerSerTyrHisAlaHisProGlnLysMetAsnPheValAlaPro 405
Db 2410 CCCTTCCCGCGCTCTCC-----AACTCAACTCATGCGC--- 2445
QY 406 HisProProAlaLeuProValThrSerSerSerPhePheAlaAlaProAsnProTyrTrp 425
Db 2446 -----GCCTCGCTGGAGGTGCGCGCGCGCT----- 2472
QY 426 AsnSerProThrGly 430

```

Db 2473 ---TCTCTACTGGC 2484

RESULT 10

US-08-469-412A-1

; Sequence 1, Application US/08469412A

; Patent No. 5856125

; GENERAL INFORMATION:

; APPLICANT: Mavrothalassitis, George J.

; APPLICANT: Blair, Donald G.

; APPLICANT: Fisher, Robert J.

; APPLICANT: Beal Jr., Gregory J.

; APPLICANT: Athanasios, Meropi A.

; APPLICANT: Sgouras, Dionysios N.

; TITLE OF INVENTION: The ERF Genetic Locus and Its Products

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/469,412A

; FILING DATE: 05-JUN-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Garrett-Wackowski, Eugenia

; REGISTRATION NUMBER: 37,330

; REFERENCE/DOCKET NUMBER: 015280-229000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2667 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 123..1769

; OTHER INFORMATION: /note= "human ERF (ETS2 Repressor

; OTHER INFORMATION: Factor) CDNA"

US-08-469-412A-1

Alignment Scores:

Pred. No.:	5,72e-26	Length:	2667
Score:	342.50	Matches:	84
Percent Similarity:	53.40%	Conservative:	26
Best Local Similarity:	40.78%	Mismatches:	59
Query Match:	14.01%	Indels:	38
DB:	2	Gaps:	5

US-09-902-772-2 (1-451) x US-08-469-412A-1 (1-2667)

QY 239 ProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluasp--G 258

Db 74 CCGGGAGCCGGCCGGGATCGGGCGCTTCGCCCGGCCCGCCACCATGAAGACCC 133

QY 258 lAArgProGlnLeuaspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAnP 278

Db 134 GCGGACACAGAGGTTGCTTCGGGAT---GGCCCTACA---AGCCAGAG-TGCTGCC 186

QY 278 rGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuLeuSerSerSera 298

Db 187 CTGGCTCAGGAGATCCAGCTGTGGCATTATCTCGAGCTGCTGGGAGGAGGT 246

QY 298 snSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAspG 318

Db 247 ACCAGGCGCTCATTCGCTGGCAGGGGAGTTCAGGGGAATTCGTCATCAAGACCCCTGATG 306

QY 318 luValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuS 338

Db 307 AGGTGGCCCGGCTGTGGGGCGTTCGCAAGTGCAGCCCGCAGATGATACGCAAGCTGA 366

QY 338 erArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgT 358

Db 367 GCGGGCGCTGCGCTATTACTATACAGCGCATTCGCACAGACCAAGGGGAAACGCT 426

QY 358 yrAlaTyrLysPheAspPheHis----- 365

Db 427 TCACCTACAAAGTTCAATTTCAACAAACTGGTCTGGTCAATTACCATTCATTGATGTGG 486

QY 366 -----GlyIleAlaGlnAlaLeuGlnProHisProProGlnSerSerMetT 381

Db 487 GGTGGCTGGGGTGCAGTGGCCCCAGAGTGGCCCCGAGTGGCGTGGGTGGTAGCCT 546

QY 381 yrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetA 401

Db 547 TCGCTTCCTCCCTCAAGCCCC-----HisProProAlaLeuProValThrSerSerS 416

QY 401 snPheValAlaPro----- 416

Db 574 AGTGTGTGCCCCCAGCGAGGACCCCGCTCACCACCGAGCTGCTCTTCATCTTCATCTT 633

QY 416 erPhePheAlaAla 420

Db 634 CCTCTTCTCGGCT 647

RESULT 11

US-09-021-715-1

; Sequence 1, Application US/09021715

; Patent No. 6194547

; GENERAL INFORMATION:

; APPLICANT: Mavrothalassitis, George J.

; APPLICANT: Blair, Donald G.

; APPLICANT: Fisher, Robert J.

; APPLICANT: Beal Jr., Gregory J.

; APPLICANT: Athanasios, Meropi A.

; APPLICANT: Sgouras, Dionysios N.

; TITLE OF INVENTION: The ERF Genetic Locus and Its Products

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/021,715

; FILING DATE: 10-Feb-1998

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Garrett-Wackowski, Eugenia

; REGISTRATION NUMBER: 37,330

; REFERENCE/DOCKET NUMBER: 015280-229000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2667 base pairs

; TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 123..1769
OTHER INFORMATION: /note= "human ERF (ETS2 Repressor
Factor) CDNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-021-715-1
Alignment Scores:
Pred. No.: 5,72e-26 Length: 2667
Score: 342.50 Matches: 84
Percent Similarity: 53.40% Conservative: 26
Best Local Similarity: 40.78% Mismatches: 59
Query Match: 14.01% Indels: 38
DB: 3 Gaps: 5
US-09-902-772-2 (1-451) x US-09-021-715-1 (1-2667)
QY 239 ProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp--G 258
DB 74 CCGGGAGCCCGCGCGCGGATCGGGCGCTTCGCCCGGGCCCCCAGCATGAAGACCC 133
QY 258 InArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 278
DB 134 GCGGACACAGGGTTTCCCTCCCGATT--GGGCTACA---AGCCAGAG--TCGTCCC 186
QY 278 roGlySerGlyGlnIleGlnLeuThrGlnPheLeuLeuGluLeuLeuSerSerSerA 298
DB 187 CTGGCTCAAGGACATCCAGCTGGCATTATCTCTGGAGCTGCGGAGAGGAGT 246
QY 298 snSerAsnCyseTheThrTriGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 318
DB 247 ACAGGGCGTCATTCCTGCGAGGGGACTACGGGAATTCGTCATCAAGACCCCTGATG 306
QY 318 luValAlaAgaTgTgTgGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 338
DB 307 AGGTGGCCCGCTGCGGGCGTTCGCAAGTGCAAGCCCGAGATGAATACGACAGCTGA 366
QY 338 erArgAlaLeuAgaTgTgTgTyrAspLysAsnIleMetThrLysValHisGlyLysArg 358
DB 367 GCCGGCCCTCGCTATCTATACCAAGCGCATTCGCACAGACCCAGGGAACGGT 426
QY 366 -----GlyIleAlaGlnAlaLeuGlnProHisProProGluSerSerMet 381
DB 487 GGTGGCTGGGGTGCAGTGCCTCCAGAGTGCCTCCAGTGCCTGGGTGGTACCACT 546
QY 381 yrlYstYrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 401
DB 547 TCCGCTTCCTCCCTCAAGGCC-----TCCG 573
QY 401 snPheValAlaPro-----HisProProAlaLeuProValThrSerSer 416
DB 574 AGGTGCTGCTCCCGCAGGAGCCCGCTCACCACGAGCCTGCTCTTCATCTTCATCTT 633
QY 416 erPhePheAlaAla 420
DB 634 CCCCTCTCTCGCT 647
RESULT 12
US-08-875-944B-1
; Sequence 1, Application US/08875944B
; Patent No. 6096542
; GENERAL INFORMATION:
; APPLICANT: FUJINAGA, Kei
; APPLICANT: YOSHIDA, Koichi
; APPLICANT: HIGASHINO, Fumihiko

TITLE OF INVENTION: CANCER CONTROL
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 624 Ninth Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,944B
FILING DATE: 07-AUG-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 07-020173
FILING DATE: 08-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/00016
FILING DATE: 09-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, ROGER L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: FUJINAGA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2664 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1386
US-08-875-944B-1
Alignment Scores:
Pred. No.: 5,62e-24 Length: 2064
Score: 322.00 Matches: 129
Percent Similarity: 38.41% Conservative: 50
Best Local Similarity: 27.68% Mismatches: 152
Query Match: 13.18% Indels: 137
DB: 3 Gaps: 17
US-09-902-772-2 (1-451) x US-08-875-944B-1 (1-2064)
QY 45 SerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProProAlaArgVal 64
DB 218 TCAAG 247
QY 65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84
DB 248 -----ACCGGCCCTGTCTGTCAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
QY 85 CysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyr 104
DB 290 -----ACCATGGGAG 307
QY 105 GlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArg 124
DB 308 -----TTTACTCCAGTGCCTATGACCCCCCAGACAAATCGCATCAAGTCCCTTACC 358
QY 125 ValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGlnTyrLeuGlu 144
DB 359 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 394

145 TrpAlaValIysGluTyrGly-----LeuProAspValAspIleLeuLeu 159
: : : : :
395 TTCCCGGGCAGACGAAATTCCTGAGATCCTCTGGCACCTCCAGCCGCCACCTG 454
160 PheGlnAsnIleAspGlyLeuLeuCysLysMetThrLysAspPheGlnArgLeu 179
: : : : :
455 GCCATGGGTACTCGGG-----AACATA
180 ThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGly 199
: : : : :
479 GCTCGGTCTCCAGACGCCCTCGACATTCCTGACTCTTCACATCTCAGGAGGGGGCC 538
200 AlaThrPheIlePhePro-----AsnThrSer-----ValTyrProGluAla 213
: : : : :
539 GGGAAACCCCTCCAGCCCTTACCAACACAGCTGTGGAGCCCTGCCACCTATCCCC 598
214 ThrGlnArgIleThrArgProAspLeuProTyrGluGlnAlaArgAsp----- 231
: : : : :
599 AGCAGAGCTTTAAGCAAGAAATACCATGATCCCTGTATGAACAGCGCGGCCAGCCG 658
232 -----AlaTrpThrSerHisSer 237
659 TGGACCAAGGTGGGTCAATGGCACAGGTACCCAGGGCGGGGTGGTGTATCAACAGG 718
238 HisProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp 257
: : : : :
719 AACAGACGGAATTCGCTACAGCTCAGATGTCACGGGTGGCATCAATGTACTCCACA 778
258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270
: : : : :
779 CAGAGGGTCTCTGGGCCCTCTCCAGGTGACGGGCCATGGGTATGGCTATGAGAAAC 838
271 -----ThrSer 272
839 CTCTGCGACCATCCAGATGATGTCTGCTGTCTCCGTGAGAAATTTGAAGGAGACATCA 898
273 SerArgLeuAlaAsnProGly-----SerGlyGlnIle 283
: : : : :
899 AGCAGGAAGGGTGGTGCATTCGAGAGGGCGGCCCTACAGCGCGGGGTGCCCTG 957
284 GlnLeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCysIleThr 303
: : : : :
958 CAGCTGTGGCAATTTCTGGTGGCTTGTGATGACCCACCAAAATGCCCATTTCAATGCC 1017
304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgTyr 323
: : : : :
1018 TGGACGGCCGGGAATCGATTCAGCTCAATGAGCTGAGGCTGAGGCTGCCAGCTCTGG 1077
324 GlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerAlaLeuArgTyr 343
: : : : :
1078 GGCATCCAGAAAGACCGGCCAGCATGAATACGACAGCTGAGCGCGCTGCTCCGATAC 1137
344 TyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAsp 363
: : : : :
1138 TATTATGAGAAGGCATCATCAGAAAGTGGCTGGTGAAGCTTACGTGTACAAAGTTT--- 1194
364 PheHisGlyIleAlaGlnAla-LeuGlnProHisProGluSerSerMetTyrLysTyr 383
: : : : :
1195 -----GTGTGTGAGCCGAGGCC-----TCTTCTCTT-----TG 1224
383 rProSer-----AspLeuProTyrMetSerSerTyrHisAlaHisProGly 398
: : : : :
1225 GCCTTCCCGCAATACAGCTCCAGCTCTCAAGGTGAGTTTGACCGCCCTGTCCAGTGAG 1284
398 nLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPheP 418
: : : : :
1285 GAGGACAGTCCCTTTGTCCTCCACTTGGATGAGACCCCGCTACTTCCAGAGCTGGCT 1344
418 eAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyTyr-----ProAsnTh 436
: : : : :
1345 GSCCCCGCCAG-CCATTT-----GGCCCCAAGGGTGGCTACTTACTAGCCCCAGC 1397
436 rArgLeuProAlaAla 441

Db 1398 GGCTGTTCCTCCCTGGC 1413
RESULT 13
US-09-116-049-3
; Sequence 3, Application US/09116049A
; Patent No. 6248351
; GENERAL INFORMATION:
; APPLICANT: Hung, Men-Chie
; TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
; FILE REFERENCE: UTSC:582
; CURRENT APPLICATION NUMBER: US/09/116, 049A
; CURRENT FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-116-049-3
Alignment Scores: 5.62e-24 Length: 2064
Pred. No.: 322.00 Matches: 129
Score: 322.00
Percent Similarity: 38.41% Conservative: 50
Best Local Similarity: 27.68% Mismatches: 152
Query Match: 13.18% Indels: 137
DB: 17
US-09-902-772-2 (1-451) x US-09-116-049-3 (1-2064)
QY 45 SerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArgVal 64
218 TCAGAGAGAGGCCAGATCCCGCACAG----- 247
65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84
248 -----ACCGGGCCCTGCTCTCGACGAGGAGGAGCGGCACCTCCCTAC----- 289
85 CysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyr 104
290 -----ACCATGGGACAGTGCC----- 307
105 GlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgArg 124
308 -----TTTACTCCAGTGCCTATGACCCCGCCAGACAAATCGCCATCAAGTCCCTG 358
125 ValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGlu 144
359 CCCTCGTGGTCCCTTGGACAGT-----CGCCCTACAGCCCT 394
145 TrpAlaValIysGluTyrGly-----LeuProAspValAspIleLeuLeu 159
395 TTCCCGGGCAGACGAAATTCCTGAGATCCTCTGGCACCTCCAGCCGCCACCTG 454
160 PheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlnArgLeu 179
455 GCCATGGGTACTCGGG-----AACATA 478
180 ThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGly 199
479 GCTCGGTCTCCAGACGCCCTCGACATTCCTGACTCTTCACATCTCAGGAGGGGGCC 538
200 AlaThrPheIlePhePro-----AsnThrSer-----ValTyrProGluAla 213
539 GGGAAACCCCTCCAGCCCTTACCAACACAGCTGTGGAGCCCTGCCACCTATCCCC 598
214 ThrGlnArgIleThrArgProAspLeuProTyrGluGlnAlaArgAsp----- 231
599 AGCAGAGCTTTAAGCAAGAAATACCATGATCCCTGTATGAACAGCGCGGCCAGCCG 658
232 -----AlaTrpThrSerHisSer 237

Db	659	TGGACCAAGGTGGGGTCAATGGGCACAGGTACCCAGGGGGGGGGTGGTGATCAACACAGG	718
Qy	238	HisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAsp	257
Db	719	AACAGACGGACTTCGGCTACGACTCAGACTGCACCGGGTGGCATCAATGTACCTCCACA	778
Qy	258	GlnArgProGlnLeuAspProTyThrGlnIleLeuGlyPro	270
Db	779	CAGAGGGCTTCTCTGGGCCCTCTCCAGGTGACGGGGCCATGGGCTATGGCTATGAGAAC	838
Qy	271	-----ThrSer	272
Db	839	CTCTGGACACATCCCCAGATGATGTCTGGGTGTCTCTGAGAAATTTGAAGGAGACATCA	898
Qy	273	SerArgLeuAlaAsnProGly-----SerGlyGlnIle	283
Db	899	AGCAGGAAGGGTCCGT-GCATTTCCGAGGGGGCGCCCTACCAGCGCCGGGTGCCCTG	957
Qy	284	GlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThr	303
Db	958	CAGCTGTGGCAATTCGTGGTGGCCCTCTGTGATGACCCCAAGAGCCCATTCATTCATGCC	1017
Qy	304	TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaAlaArgArgTyr	323
Db	1018	TGGACGGCGCGGGAATCGAGTTCAGCTCATTTGAGCCTGAGGAGGTGCGCAGGCTCTGG	1077
Qy	324	GlyGluArgLysSerLysProAsnMetAsnTyrAspLysLysLeuSerArgAlaLeuArgTyr	343
Db	1078	GGCATCCAGAAGAACCGGCCAGCCATCAATTACACAAAGCTGAGCGCTCGCTCCGATAC	1137
Qy	344	TyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAsp	363
Db	1138	TATTATGAGAAAGGCATCATGCAGAGGTGGCTGGTGAAGCTTACGTGTACAGTTT---	1194
Qy	364	PheHisGlyIleAlaGlnAla-LeuGlnProHisProGluSerSerMetTyrLysTyr	383
Db	1195	-----GTGTGTGAGCCCGAGGCC-----TCTTCTCTT-----TG	1224
Qy	383	rProSer-----AspLeuProTyrMetSerSerTyrHisAlaHisProGln	398
Db	1225	GCCTTCGGCAATACAGCGTCCAGCTCTCAAGGCTGAGTTTGACGGCGCTGTGAGTGAG	1284
Qy	398	nLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhePh	418
Db	1285	GAGGACACAGTCCCTTTGTCCCATTTGGATGAGAGCCCGGCTACTCTCCAGAGCTGGCT	1344
Qy	418	eAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyr-----ProAsnThr	436
Db	1345	GGCCCCGCCAG-CAATT-----GGCCCCAAGGGTGGCTACTCTTAGCCCCCAGC	1397
Qy	436	rArgLeuProAlaAla	441
Db	1398	GGCTGTCCCCCTGCC	1413

RESULT 14

US-09-602-868A-1

US-09-802-808A-1
; Sequence 1, Application US/09602868A

; sequence 1, 500110
; Patent No. 6528310

GENERAL INFORMATION:

APPLICANT: FUJINAGA, Kei

APPLICANT: YOSHIDA, Koichi

APPLICANT: HIGASHINO, Fumihiro

1. TITLE OF INVENTION: CANCER CONTROL

; NUMBER OF SEQUENCES: 5
CONSEQUENCE ADDRESS:

```

; CORRESPONDENCE ADDRESS:
; ADDRESS, BPOBY AN

```

ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 624 Ninth Street N.W. Ste. 300

STREET: 624 NINTH STREET N.W., Ste. 300
CITY: Washington

CITY: Washington
STATE: D.C.STATE: D.C.
COUNTRY: USA

COUNTRY: USA
ZIP: 20001

REF. 2001
COMPUTER READAB

MEDIUM TYPE: Floppy

1000

```
Qy 200 AlaThrPheIlePhePro-----AsnThrSer-----ValTyrProGluAla 213
Db 539 GGGAAACCCCTCCAGCCCTTACCAACACACCGCTGTGGAGCCCTGCCACCTTATCCCC 598
Qy 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgAgsSer----- 231
Db 599 AGCAGAGCTTTTAAGCAAGATAACATGATCCCTGTATGAACAGCGCGGCAGCCAGCCG 658
Qy 232 -----AlaThrThrSerHisSer 237
Db 659 TGGACACAGGTGGGTCAATGGGCACAGTACCCAGCGCGGGGGTGGTATCAACAGG 718
Qy 238 HisProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp 257
Db 719 AACAGACGACTTCGCTTACGACTCAGATGTCACGGGTGGCATCAATGTACCTCCACA 778
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270
Db 779 CAGAGGGCTTCTCTGGGCCCTCTCCAGGTGACGGGGCCATGGGCTATGGCTATGAGAAAC 838
Qy 271 -----ThrSer 272
Db 839 CTCCTCGACCATTCACAGATGATGCTGCTGTTGCTCCCTGAGAAATTTGAAGGACATCA 898
Qy 273 SerArgLeuAlaAsnProGly-----SerglyGlnIle 283
Db 899 AGCAGAAAGGGTGGT-GCATTTCGAGAGGGCGGCCCTACCCAGCGCGGGGTGCGCTG 957
Qy 284 GlnLeuThrGlnPheLeuLeuLeuLeuSerAspSerSerSerSerSerSerSerSerSer 303
Db 958 CAGCTGTGGCAATTTCTGGTGGCTTCTGGATGATGATGATGATGATGATGATGATGATG 1017
Qy 304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaAlaArgArgTyr 323
Db 1018 TGGACGGCGCGGGAATGGAGTCAAGCTCATTCAGCTGAGCTGAGAGGTGCGCAGGCTCTG 1077
Qy 324 GlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyr 343
Db 1078 GGCATCCAGAAAGGCAATCATCAGAGGTGCTGCTGAGCTGAGAGGTGAGAGGTGAGAGGT 1137
Qy 344 TyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAsp 363
Db 1138 TATTATGAGAAAGGCAATCATCAGAGGTGCTGCTGAGCTGAGAGGTGAGAGGTGAGAGGT 1194
Qy 364 PheHisGlyIleAlaGlnAla-LeuGlnProHisProProGluSerSerMetTyrLysTy 383
Db 1195 -----GTGTGTGAGCCCGAGGCC-----TCTTCTCTT-----TG 1224
Qy 383 rProSer-----AspLeuProTyrMetSerSerTyrHisAlaHisProGly 398
Db 1225 GCCTTCCGGACAATCAGCGTCCAGCTCTCAGGCTGAGTTGACCGGCTGTGAGTGAG 1284
Qy 398 nLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPheP 418
Db 1285 GAGGACAGAGTCCCTTTGCTCCACTTGGATGAGAGCCCGGCTACCTCCAGAGCTGGCT 1344
Qy 418 eAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyr-----ProAsnTh 436
Db 1345 GCCTCCCGCCAG-CCATTT-----GGCCCCAAGGGTGGCTACTCTTACTAGCCCCCAGC 1397
Qy 436 rArgLeuProAlaAla 441
Db 1398 GCCTGTTCCTCCCTGCC 1413
```

RESULT 15

```
US-09-884-363-3
; Sequence 3, Application US/09884363
; Patent No. 6582725
; GENERAL INFORMATION:
; APPLICANT: Hung, Men-Chie
; TITLE OF INVENTION: HUMAN PE3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
; FILE REFERENCE: UTSC:562
```

```
; CURRENT APPLICATION NUMBER: US/09/884,363
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 09/116,049
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-884-363-3

Alignment Scores:
Pred. No.: 5,62e-24 Length: 2064
Score: 322.00 Matches: 129
Percent Similarity: 38.41% Conservative: 50
Best Local Similarity: 27.68% Mismatches: 152
Query Match: 13.18% Indels: 137
DB: Gaps: 17

US-09-902-772-2 (1-451) x US-09-884-363-3 (1-2064)

Qy 45 SerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProProAlaArgVal 64
Db 218 TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 247
Qy 65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84
Db 248 -----ACCCGGCCCTGCTCTGCAGCAGGAGAGCGCCACTCCCTACCC 289
Qy 85 CysSerValAlaLysGlyGlyLysMetValSerSerSerSerSerSerSerSerSer 104
Db 290 -----ACCATGGCGAGAGTCCC 307
Qy 105 GlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgArg 124
Db 308 -----TTTACTCCAGTGCCTATGACCCCGCCAGACAAATCGCCATCGCCCTG 358
Qy 125 ValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGlnTrpLeuGlu 144
Db 359 CCCTGTGCTCCCTTGACAGT-----CGCCCCATACAGCCCT 394
Qy 145 TrpAlaValLysGluTyrGly-----LeuProAspValAspIleLeuLeu 159
Db 395 TTCCTCCGGCAGAGCAACGGAATTTCTGAGATCTCTGCGACCTCCCGACCCCGCCCTG 454
Qy 160 PheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeu 179
Db 455 GCATGGGTACCTCGGG-----AACATA 478
Qy 180 ThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGly 199
Db 479 GCTCCGTCTTCCAGCAGCCCTGGACATTTGACATCTTCCATCTCAGGAGGGGGCC 538
Qy 200 AlaThrPheIlePhePro-----AsnThrSer-----ValTyrProGluAla 213
Db 539 GGGAAACCCCTCCAGCCCTTACCAACACACCGCTGTGGAGCCCTGCCACCTTATCCCC 598
Qy 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgAgsSer----- 231
Db 599 AGCAGAGCTTTTAAGCAAGATAACATGATCCCTGTATGAACAGCGCGGCAGCCAGCCG 658
Qy 232 -----AlaThrThrSerHisSer 237
Db 659 TGGACACAGGTGGGTCAATGGGCACAGTACCCAGCGCGGGGGTGGTATCAACAGG 718
Qy 238 HisProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp 257
Db 719 AACAGACGACTTCGCTTACGACTCAGATGTCACGGGTGGCATCAATGTACCTCCACA 778
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270
Db 779 CAGAGGGCTTCTCTGGGCCCTCTCCAGGTGACGGGGCCATGGGCTATGGCTATGAGAAAC 838
```

```
QY 271 -----ThrSer 272
Db 839 CTCTGCGACCATCCCGATGATCTCGGTTCCTCGAGAAATTTGAAGGAGACATCA 898
QY 273 SerArgLeuAlaAsnProGly-----SerGlyGlnIle 283
Db 899 AGCAGGAAGGGTCCGT-GCATTTCGAGAGGGCGCCCTACCAGCGCCGGGTGCCCTG 957
QY 284 GlnLeuTrpGlnPheLeuLeuLeuSerSerSerSerSerSerSerSerSerSerSer 303
Db 958 CAGCTGTGGCAATTCTGGTGGCTTGTGATGACCCAAACAAATGCCCATTTCAATGCC 1017
QY 304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgTyr 323
Db 1018 TGGACGGCCGGGAATGAGTTCAGCTCATTGAGCTGAGGAGGTCCGACGGCTCTGG 1077
QY 324 GlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyr 343
Db 1078 GGCATCCAGAAAGACCCGCCAGCCATGAATTACGACAAAGCTGAGCGCTCGCTCCGATAC 1137
QY 344 TyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAsp 363
Db 1138 TATTATGAAAGGCAATGACAGAGGTGCTGCTGAGCGTTACGTGTACAGTTT--- 1194
QY 364 PheHisGlyIleAlaGlnAla-LeuGlnProHisProHisProGluSerSerMetTyrLysTyr 383
Db 1195 -----GTGTGTGAGCCCGAGGCC-----TCTTCTCTT-----TG 1224
QY 383 rProSer-----AspLeuProTyrMetSerSerTyrHisAlaHisProGlu 398
Db 1225 GCCTTCCCGGACATCAGCTCCAGCTCTCAGGCTGAGTTGACCGGCTGTCCAGTGAG 1284
QY 398 nLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePh 418
Db 1285 GAGGACACAGTCCCTTTGTCCCACTTGATGAGAGCCCGCCTACCTCCAGAGCTGGCT 1344
QY 418 eAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyr-----ProAsnTh 436
Db 1345 GGCCCCGGCCAG-CAATTT-----GGCCCCAAGGTGGGTACTCTTACTAGCCCCCAGC 1397
QY 436 rArgLeuProAlaAla 441
Db 1398 GGCTGTTCCTCCCTGCC 1413
```

Search completed: May 27, 2004, 23:58:56
Job time : 149 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 27, 2004, 22:55:58 ; Search time 668 Seconds
(without alignments)
3071.189 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTIKKALSVSDQSLPE.....IYENTRLPRAHPMSHLGTYV 451

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0902772/runat_27052004_162843_2259/app_query.fasta_1.647
-DB_Published Applications NA -OPT=fastp -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0902772@cgn_1_1_481@runat_27052004_162843_2259
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1					

1	2444	100.0	1447	9	US-09-902-772-1	Sequence 1, Appli
2	2414.5	98.8	1528	9 <td>US-09-902-772-3</td> <td>Sequence 3, Appli</td>	US-09-902-772-3	Sequence 3, Appli
3	2182	89.3	1389	13	US-10-211-462-41	Sequence 41, Appli
4	2182	89.3	1389	13	US-10-087-192-1127	Sequence 1127, Ap
5	2182	89.3	1366	15	US-10-205-823-98	Sequence 98, Appli
6	2182	89.3	1366	15	US-10-021-660-26	Sequence 26, Appli
7	2057	84.2	1297	13	US-10-087-192-1124	Sequence 1124, Ap
8	1624.5	66.5	1359	12	US-10-052-482-198	Sequence 198, App
9	1624.5	66.5	1359	12	US-10-052-482-197	Sequence 197, App
10	1624.5	66.5	2957	13	US-10-342-887-710	Sequence 710, App
11	1624.5	66.5	2957	13	US-10-172-118-710	Sequence 710, App
12	1624.5	66.5	2957	15	US-10-007-926A-295	Sequence 295, App
13	1613.5	66.0	1359	12	US-10-052-482-194	Sequence 194, App
14	1613.5	66.0	1729	12	US-10-052-482-194	Sequence 194, App
15	956.5	39.1	21231	13	US-10-087-192-1126	Sequence 1126, Ap
16	949	38.8	69090	13	US-10-087-192-1123	Sequence 1123, Ap
17	948	38.8	567	9	US-09-864-761-18410	Sequence 18410, A
18	863	35.3	473	9	US-09-864-761-20472	Sequence 20472, A
19	857	35.1	473	15	US-10-029-386-25224	Sequence 25224, A
20	833.5	34.1	73725	12	US-10-052-482-196	Sequence 196, App
21	813.5	33.3	72732	12	US-10-052-482-193	Sequence 193, App
22	712	29.1	549	9	US-09-923-779-54	Sequence 54, Appli
23	650	26.6	420	9	US-09-864-761-3705	Sequence 3705, Ap
24	650	26.6	454	9	US-09-864-761-1853	Sequence 1853, Ap
25	638	26.1	467	10	US-09-918-995-23356	Sequence 23356, A
26	511	20.9	1752	9	US-09-850-799-1	Sequence 1, Appli
27	511	20.9	1752	15	US-10-027-859-1	Sequence 1, Appli
28	453.5	18.6	380	15	US-10-263-828-41	Sequence 41, Appli
29	451	18.5	1496	13	US-10-087-192-344	Sequence 344, App
30	449	18.4	1450	16	US-10-151-937-126	Sequence 126, App
31	449	18.4	1602	13	US-10-087-192-347	Sequence 347, App
32	446	18.2	4991	12	US-10-152-319A-1502	Sequence 1502, Ap
33	445.5	18.2	2269	9	US-09-954-531-955	Sequence 955, App
34	444.5	18.2	2188	9	US-09-920-300A-1716	Sequence 1716, Ap
35	444.5	18.2	2188	14	US-10-033-528-1716	Sequence 1716, Ap
36	444.5	18.2	2188	15	US-10-039-926-1716	Sequence 1716, Ap
37	444.5	18.2	2268	9	US-09-920-300A-1693	Sequence 1693, Ap
38	444.5	18.2	2268	14	US-10-033-528-1693	Sequence 1693, Ap
39	444.5	18.2	2268	15	US-10-039-926-1693	Sequence 1693, Ap
40	444.5	18.2	3692	15	US-10-106-698-824	Sequence 824, App
41	441.5	18.1	2777	16	US-10-108-260A-1528	Sequence 1528, Ap
42	431.5	17.7	1884	9	US-09-925-300-420	Sequence 420, App
43	407	16.7	2532	15	US-10-037-270-869	Sequence 869, App
44	407	16.7	2532	16	US-10-117-722-869	Sequence 869, App
45	401.5	16.4	3695	13	US-09-805-020-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-09-902-772-1
; Sequence 1, Application US/09902772
; Patent No. US20020164739A1
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; TITLE OF INVENTION: the Proteins
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/09/902,772
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US/08/878,177
; PRIOR FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA
US-09-902-772-1

Alignment Scores: 3.57e-269 Length: 1447
Pred. No.: 1

```
Score: 2444.00 Matches: 451
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-902-772-2 (1-451) x US-09-902-772-1 (1-1447)
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
DB 63 ATGGCAAGCACTATTAGGAAGCATTTATCAGTGGTGAAGCAAGCCAGTCTTTGTTGAG 122
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
DB 123 TGTGCTACGATCGCCCACTTTCAGAGACAGAAATGACAGCCCTCTCTCCAGTGAA 182
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
DB 183 TATGGCAACATCAAGATGAGCCCGCGTTCACCAAGAGAGAGAGAGAGAGAGAGAG 242
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 243 CCGGCCAGAGTTACATTAAGATGAGTGAACCCAAACCCAGGTTAATGGGTCAAGGAAT 302
QY 81 SerProAspCysSerValAlaLysGlyGlyLeuMetValSerSerSerAspAsnVal 100
DB 303 TCACCTGATGACTGCGGTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisLeuProProAsnMetThrThr 120
DB 363 GGGATGAACATATGGAAGCTACATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
QY 121 AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
DB 423 AATGAACGAGAGATTATGTCAGCAGATCTTACGTATGAGAGAGAGAGAGAGAGAGAG 482
QY 141 GlnTrpLeuGlnTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
DB 483 CAGTGGCTGGAGTGGGAGTGAAGAGATGATGCTCTCCAGAGCTGGAGATCTTGTGTT 542
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
DB 543 CAGAACATTGATGGGAAGAGTTGTGTAATGATCCAAAGATGACTTCCAGAGACTCAG 602
QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAla 200
DB 603 CCGAGCTATACGAGAGATATCTCTGTCCACCTACACTACCTCAGAGAGAGAGAGAG 662
QY 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220
DB 663 ACTTTTATTTTCCAAATATACATCAGTTTACCCAGAGCAACGCAAGAGATAACCAAGG 722
QY 221 ProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThr 240
DB 723 CCAGATTACCTTATGAGCAAGCAGAGAGATCAGCGTGGAGAGTCAACCCATCCCACT 782
QY 241 GlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAspGlnArgPro 260
DB 783 CAGTCAAAAGCTACCAACCATCATCTTCAACAGTGCACCAAGAGAGAGAGAGAGAGAG 842
QY 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
DB 843 CAGTTAGATCTTATCAGATTCTTGGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
QY 281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsn 300
DB 903 GGGCAGATACAGCTATGSCAGTTCTTCTGAGGCTTCTGCGGAGAGCTCCAACTCCAAC 962
QY 301 CysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320
DB 963 TSCATCACCTGGGAGGGCACAATAATGGGAGATCAAGATGACAGACCTGATGAAGTGCT 1022
QY 321 ArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
```

RESULT 2

US-09-902-772-3

; Sequence 3, Application US/09902772

; Patent No. US20020164739A1

; GENERAL INFORMATION:

; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et

; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of

; TITLE OF INVENTION: the Proteins

; FILE REFERENCE: chugai selyaku kabushiki kaisha 5001

; CURRENT APPLICATION NUMBER: US/09/902,772

; PRIOR FILING DATE: 2001-07-12

; CURRENT APPLICATION NUMBER: US/08/878,177

; PRIOR FILING DATE: 1997-06-18

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 1528

; TYPE: DNA

; ORGANISM: c-erg gene, chicken DNA

US-09-902-772-3

Alignment Scores:

Pred. No.:	9,186-286	Length:	1528
Score:	2414.50	Matches:	450
Percent Similarity:	94.14%	Conservative:	0
Best Local Similarity:	94.14%	Mismatches:	1
Query Match:	98.79%	Indels:	27
DB:	9	Gaps:	1

US-09-902-772-2 (1-451) x US-09-902-772-3 (1-1528)

QY	1	MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu	20
DB	63	ATGGCAAGCACTATTAGGAAGCATTTATCAGTGGTGAAGCAAGCCAGTCTTTGTTGAG	122
QY	21	CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu	40
DB	123	TGTGCTACGATCGCCCACTTTCAGAGACAGAAATGACAGCCCTCTCTCCAGTGAA	182
QY	41	TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro	60
DB	183	TATGGCAACATCAAGATGAGCCCGCGTTCACCAAGAGAGAGAGAGAGAGAGAGAG	242
QY	61	ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn	80
DB	243	CCGGCCAGAGTTACATTAAGATGAGTGAACCCAAACCCAGGTTAATGGGTCAAGGAAT	302

QY 81 SerProAspCysSerValAlaLysGlyLysMetValSerSerSerAspVal 100
Db 303 TCACCTGATGACTGACGGTGGCAAGAGGAGGAAATGTTAGCAGTTTCAGACATGTT 362
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisLeuProProAsnMetThrThr 120
Db 363 GGGATGAACATATGGAAGCTACATGGAAGAGGAGCATATTCGGCTCCAAATATGACAAAC 422
QY 121 AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
Db 423 AATGAACGAAGAGTTATTTGCCAGCAGATCTTACGTTNTGGAGCAGACACCATGTACGG 482
QY 141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
Db 483 CAGTGGCTGGAGTGGGAGTGAAGAGTATGGTCTTCAGACGGTGGACATCTTTGTTGTC 542
QY 161 GlnAsnIleAspGlyLysGluLeuLysMetThrLysAspPheGlnArgLeuThr 180
Db 543 CAGAACATTTGATGGGAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGACTCAG 602
QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg----- 198
Db 603 CCGAGCTATAACGAGATATCTCTGTACACCTACACTACCTCAGAGACTCCTCTT 662
QY 198 ----- 198
Db 663 CCACATTTGACTTCAGATGATGTTGATAAGGCTTACAAACTCTCCACGGTTAATGCAT 722
QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213
Db 723 GCTAGAAACACAGAGGAGCCACTTTATTTTCCAAATACATCAGTTTACCAGAGCA 782
QY 214 ThrGlnArgIleThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrp 233
Db 783 ACGCAAGAAATAACAACAGCCAGATTTACCTTATGAGCAAGCGAGAGATCAGCGTGG 842
QY 234 ThrSerHisSerHisProThrGlnSerLysIleThrGlnProSerSerSerThrValPro 253
Db 843 ACGAGTCACACCATCCACATCAGTCAAAAGCTACCAACCATCATCTTCAACAGTGCCT 902
QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273
Db 903 AAACAGAGAGACCGCTCTCAGTAGATCTCTTATCAGATCTTGGACCGACGAGCAGC 962
QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeu 293
Db 963 CGTCTTGCAAAATCCAGGAGTGGGAGATACAGTATGCGCAGTTCTTACTGGAGCTTCTG 1022
QY 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
Db 1023 TCGNACAGCTCCACTCCAACTGCATCCTGGAGGGCACAATGGGAGTTCAAGATG 1082
QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333
Db 1083 ACAGACCTGTATGAATGGCTCGCGGTGGGGAGAGAGAGAAAGCAAACTTAAATGAAC 1142
QY 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLysVal 353
Db 1143 TATGCAAACTCAGCGGTGCATCTCGCTACTACTATGACAAATATATTATGACTTAAATG 1202
QY 354 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
Db 1203 CATGGTAAACGCTATGCCTACAAATTTGATTTCCACGGAAATCGCTCAGGCGCTCCAGCCT 1262
QY 374 HisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393
Db 1263 CACCTTCAGATATCATCTGATGACAAATACCATCAGACCTCCCTCTCATGAGTTCCTAC 1322
QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413
Db 1323 CATGCACACCCACAGAGATGAACCTTTGTAGCTCCCATCCCTCTGCTTGGCCCGTAAAC 1382

QY 414 SerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyr 433
Db 1383 TCATCCAGCTTTTGTGCTGCCCTAATCCATACTGGAATTCACCAACTGGAGGCATCTAC 1442
QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1443 CCCAATACAGGCTGCCAGCTCATATGCTTCCCATCTTTGGCACCCTACTACTAC 1496
RESULT 3
US-10-211-462-41
; Sequence 41, Application US/10211462
; Publication No. US20040033495A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-211-462-41
Alignment Scores:
Pred. No.: 3,178-239 Length: 1389
Score: 2182.00 Matches: 407
Percent Similarity: 91.72% Conservative: 14
Best Local Similarity: 88.67% Mismatches: 24
Query Match: 89.28% Indels: 14
DB: 13 Gaps: 4
US-09-902-772-2 (1-451) x US-10-211-462-41 (1-1389)
QY 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
Db 25 GCAGCTCATATCAAGGAAGCTTATCAGTTGTGAGTGAGGACCAAGTCGTTGTTGAGTGT 84
QY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41
Db 85 GCCTACGGAACCGCACCTCGCTGAGCAGAGATGACCGCTCTCTCCAGCGACTAT 144
QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProPro 61
Db 145 GGCACACACTTCCAGATGAGCCACCGCTCCCTCAGCAGATTGGCTGTCTCAACCCCA 204
QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db 205 GCCAGGGTCCACCATCAAAATGGAATGTAACCTTAGCCAGGTGAATGGCTCAAGGAACCT 264
QY 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGly 101
Db 265 CTTGATGAATGAGTGTGGCCAAAGCGGGAAGATGCTGGGAGCCCGACACCGTTGG 324
QY 102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsn 121
Db 325 ATGAATACGGCAGCTACATGGAGGAGAGCAATGCCACCCCAACATGACCAAGAAC 384

122 GluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
123 |
124 |
125 |
126 |
127 |
128 |
129 |
130 |
131 |
132 |
133 |
134 |
135 GAGCGCAGAGTTATCGTGGCAGCAGATCTTACGCTATGGAGTACAGACCATGTGGCGCAG 444
136 |
137 |
138 |
139 |
140 |
141 |
142 TrpLeuGluTrpAlaValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 161
143 |
144 |
145 TGGCTGGAGTGGCGGTTGAAAGATATGGCTTCCAGACGTCACATCTTGTATTCCAG 504
146 |
147 |
148 |
149 |
150 |
151 |
152 AsnIleAspGlyLysGluLeuValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 181
153 |
154 |
155 |
156 |
157 |
158 |
159 |
160 |
161 |
162 |
163 |
164 |
165 |
166 |
167 |
168 |
169 |
170 |
171 |
172 |
173 |
174 |
175 |
176 |
177 |
178 |
179 |
180 |
181 |
182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
183 |
184 |
185 |
186 |
187 |
188 |
189 |
190 |
191 |
192 |
193 |
194 |
195 |
196 |
197 |
198 |
199 |
200 |
201 |
202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
203 |
204 |
205 |
206 |
207 |
208 |
209 |
210 |
211 |
212 |
213 |
214 |
215 |
216 |
217 |
218 |
219 |
220 |
221 |
222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTrp 233
223 |
224 |
225 |
226 |
227 |
228 |
229 |
230 |
231 |
232 |
233 |
234 ThrSerHisSerHisProThr--GlnSerIleValThrGlnProSerSerSerThrVal 252
235 |
236 |
237 |
238 |
239 |
240 |
241 |
242 |
243 |
244 |
245 |
246 |
247 |
248 |
249 |
250 |
251 |
252 |
253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272
254 |
255 |
256 |
257 |
258 |
259 |
260 |
261 |
262 |
263 |
264 |
265 |
266 |
267 |
268 |
269 |
270 |
271 |
272 |
273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 292
274 |
275 |
276 |
277 |
278 |
279 |
280 |
281 |
282 |
283 |
284 |
285 |
286 |
287 |
288 |
289 |
290 |
291 |
292 |
293 LeuSerAspSerSerAsnSerCysIleThrTrpGluGlyThrAsnGlyGluPheLys 312
294 |
295 |
296 |
297 |
298 |
299 |
300 |
301 |
302 |
303 |
304 |
305 |
306 |
307 |
308 |
309 |
310 |
311 |
312 |
313 MetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMet 332
314 |
315 |
316 |
317 |
318 |
319 |
320 |
321 |
322 |
323 |
324 |
325 |
326 |
327 |
328 |
329 |
330 |
331 |
332 |
333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLys 352
334 |
335 |
336 |
337 |
338 |
339 |
340 |
341 |
342 |
343 |
344 |
345 |
346 |
347 |
348 |
349 |
350 |
351 |
352 |
353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
354 |
355 |
356 |
357 |
358 |
359 |
360 |
361 |
362 |
363 |
364 |
365 |
366 |
367 |
368 |
369 |
370 |
371 |
372 |
373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
374 |
375 |
376 |
377 |
378 |
379 |
380 |
381 |
382 |
383 |
384 |
385 |
386 |
387 |
388 |
389 |
390 |
391 |
392 |
393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
394 |
395 |
396 |
397 |
398 |
399 |
400 |
401 |
402 |
403 |
404 |
405 |
406 |
407 |
408 |
409 |
410 |
411 |
412 |
413 ThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyIle 432
414 |
415 |
416 |
417 |
418 |
419 |
420 |
421 |
422 |
423 |
424 |
425 |
426 |
427 |
428 |
429 |
430 |
431 |
432 |
433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
434 |
435 |
436 |
437 |
438 |
439 |
440 |
441 |
442 |
443 |
444 |
445 |
446 |
447 |
448 |
449 |
450 |
451 |
452 |
453 |
454 |
455 |
456 |
457 |
458 |
459 |
460 |
461 |
462 |
463 |
464 |
465 |
466 |
467 |
468 |
469 |
470 |
471 |
472 |
473 |
474 |
475 |
476 |
477 |
478 |
479 |
480 |
481 |
482 |
483 |
484 |
485 |
486 |
487 |
488 |
489 |
490 |
491 |
492 |
493 |
494 |
495 |
496 |
497 |
498 |
499 |
500 |
501 |
502 |
503 |
504 |
505 |
506 |
507 |
508 |
509 |
510 |
511 |
512 |
513 |
514 |
515 |
516 |
517 |
518 |
519 |
520 |
521 |
522 |
523 |
524 |
525 |
526 |
527 |
528 |
529 |
530 |
531 |
532 |
533 |
534 |
535 |
536 |
537 |
538 |
539 |
540 |
541 |
542 |
543 |
544 |
545 |
546 |
547 |
548 |
549 |
550 |
551 |
552 |
553 |
554 |
555 |
556 |
557 |
558 |
559 |
560 |
561 |
562 |
563 |
564 |
565 |
566 |
567 |
568 |
569 |
570 |
571 |
572 |
573 |
574 |
575 |
576 |
577 |
578 |
579 |
580 |
581 |
582 |
583 |
584 |
585 |
586 |
587 |
588 |
589 |
590 |
591 |
592 |
593 |
594 |
595 |
596 |
597 |
598 |
599 |
600 |
601 |
602 |
603 |
604 |
605 |
606 |
607 |
608 |
609 |
610 |
611 |
612 |
613 |
614 |
615 |
616 |
617 |
618 |
619 |
620 |
621 |
622 |
623 |
624 |
625 |
626 |
627 |
628 |
629 |
630 |
631 |
632 |
633 |
634 |
635 |
636 |
637 |
638 |
639 |
640 |
641 |
642 |
643 |
644 |
645 |
646 |
647 |
648 |
649 |
650 |
651 |
652 |
653 |
654 |
655 |
656 |
657 |
658 |
659 |
660 |
661 |
662 |
663 |
664 |
665 |
666 |
667 |
668 |
669 |
670 |
671 |
672 |
673 |
674 |
675 |
676 |
677 |
678 |
679 |
680 |
681 |
682 |
683 |
684 |
685 |
686 |
687 |
688 |
689 |
690 |
691 |
692 |
693 |
694 |
695 |
696 |
697 |
698 |
699 |
700 |
701 |
702 |
703 |
704 |
705 |
706 |
707 |
708 |
709 |
710 |
711 |
712 |
713 |
714 |
715 |
716 |
717 |
718 |
719 |
720 |
721 |
722 |
723 |
724 |
725 |
726 |
727 |
728 |
729 |
730 |
731 |
732 |
733 |
734 |
735 |
736 |
737 |
738 |
739 |
740 |
741 |
742 |
743 |
744 |
745 |
746 |
747 |
748 |
749 |
750 |
751 |
752 |
753 |
754 |
755 |
756 |
757 |
758 |
759 |
760 |
761 |
762 |
763 |
764 |
765 |
766 |
767 |
768 |
769 |
770 |
771 |
772 |
773 |
774 |
775 |
776 |
777 |
778 |
779 |
780 |
781 |
782 |
783 |
784 |
785 |
786 |
787 |
788 |
789 |
790 |
791 |
792 |
793 |
794 |
795 |
796 |
797 |
798 |
799 |
800 |
801 |
802 |
803 |
804 |
805 |
806 |
807 |
808 |
809 |
810 |
811 |
812 |
813 |
814 |
815 |
816 |
817 |
818 |
819 |
820 |
821 |
822 |
823 |
824 |
825 |
826 |
827 |
828 |
829 |
830 |
831 |
832 |
833 |
834 |
835 |
836 |
837 |
838 |
839 |
840 |
841 |
842 |
843 |
844 |
845 |
846 |
847 |
848 |
849 |
850 |
851 |
852 |
853 |
854 |
855 |
856 |
857 |
858 |
859 |
860 |
861 |
862 |
863 |
864 |
865 |
866 |
867 |
868 |
869 |
870 |
871 |
872 |
873 |
874 |
875 |
876 |
877 |
878 |
879 |
880 |
881 |
882 |
883 |
884 |
885 |
886 |
887 |
888 |
889 |
890 |
891 |
892 |
893 |
894 |
895 |
896 |
897 |
898 |
899 |
900 |
901 |
902 |
903 |
904 |
905 |
906 |
907 |
908 |
909 |
910 |
911 |
912 |
913 |
914 |
915 |
916 |
917 |
918 |
919 |
920 |
921 |
922 |
923 |
924 |
925 |
926 |
927 |
928 |
929 |
930 |
931 |
932 |
933 |
934 |
935 |
936 |
937 |
938 |
939 |
940 |
941 |
942 |
943 |
944 |
945 |
946 |
947 |
948 |
949 |
950 |
951 |
952 |
953 |
954 |
955 |
956 |
957 |
958 |
959 |
960 |
961 |
962 |
963 |
964 |
965 |
966 |
967 |
968 |
969 |
970 |
971 |
972 |
973 |
974 |
975 |
976 |
977 |
978 |
979 |
980 |
981 |
982 |
983 |
984 |
985 |
986 |
987 |
988 |
989 |
990 |
991 |
992 |
993 |
994 |
995 |
996 |
997 |
998 |
999 |
1000 |

RESULT 4

US-10-087-192-1127
; Sequence 1127, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:

APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: CANCER
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1127
LENGTH: 1389
TYPE: DNA
ORGANISM: Homo sapiens
US-10-087-192-1127
Alignment Scores:
Pred. No.: 3,17e-239 Length: 1389
Score: 2182.00 Matches: 407
Percent Similarity: 91.72% Conservative: 14
Best Local Similarity: 88.57% Mismatches: 24
Query Match: 89.28% Indels: 14
DB: 13 Gaps: 4
US-09-902-772-2 (1-451) x US-10-087-192-1127 (1-1389)
QY 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
DB 25 GGAGCTCATATCAAGGAAGCCCTTATCAGTTGTGAGTGAGGACCAGTCGTTGTTGAGTGT 84
QY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41
DB 85 GCCTACGGAGACGCCACACCTGGCTAAGACAGAGATGACCGGCTCTCTCTCCAGCGACTAT 144
QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61
DB 145 GGACAGACTTCCAGATGAGCCCGCTCCCTCAGCAGGATTGGCTGTCTCAACCCCA 204
QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
DB 205 GCCAGGGTCACCATCAAAATGGAATGTAACTGACCCAGGTGAATGGCTCAAGGAACCT 264
QY 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly 101
DB 265 CCGTGAATGAGTGGCGGCAAGGGGGAAGATGGTGGCAGCCCGCCAGACCCGTTGG 324
QY 102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThrAsn 121
DB 325 ATGAACCTACGGCAGCTACATGGAGGAGAAGCACATGCCACCCCAACATGACACCAAG 384
QY 122 GluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141
DB 385 GAGCCAGAGTTATCGTCCAGCAGATCCTACGCTATGGAGTACAGACCATGTGGCGGAG 444
QY 142 TrpLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
DB 445 TGGCTGGAGTGGCGGTTGAAAGATATGGCTTCCAGACGTCACCATCTGTATTATCCAG 504
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB 505 AACATCGATGGGAAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCC 564
QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
DB 565 AGCTACACCGCGGACATCTCTCTCTCTCACAATCTCCACTACCTCAGAGAGACTCTCTTCCA 624
QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
DB 625 CATTTG-----ACTTCA-----GATGATGTTGATAAGCCTTACAAACTCTTCCA 669


```
Qy 222 -----AspLeuProTyrGluGlnAlaArgSerAlaTTP 233
Db 670 CGGTTAATGATGCTAGAAACACAGATTTTACCATAAGAGCCCGCCAGGAGATCAGCGCTGG 729
Qy 234 ThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerThrVal 252
Db 730 ACCGGTCACGGCCACCCACGCCCACTCGAAGCTGCTCAACATCTCTTCCACAGTG 789
Qy 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnLysLeuGlyProThrSer 272
Db 790 CCAAACTGAAGACCAAGCTCTCTAGTAGCTTATCAGATCTCTTGACCAACAAGT 849
Qy 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 292
Db 850 AGCGCCCTTGCATAATCCAGGAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGCTC 909
Qy 293 LeuSerAspSerSerAsnSerAsnCysIleThrTrpLysGlyThrAsnGlyGluPheLys 312
Db 910 CTGTGGACAGCTCCAACTCCAGCTGCATCACCTGGGAGGCCAACCAAGGGAGTTCAAG 969
Qy 313 MetThrAspProAspGluValAlaArgTyrGlyGluArgLysSerLysProAsnMet 332
Db 970 ATGACGGATCCCGACGAGGTGGCCGCTGGGGAGCGGAGAGCAACCAACATG 1029
Qy 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLys 352
Db 1030 AACTACGATAAGCTCAGCCGGCCCTCCGTTACTACTATGACAAAGACATCATGCCAAG 1089
Qy 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
Db 1090 GTCCATGGGAGCGCTACGGCTCAAGTTCGATTCACGGGATCCGCCAGGCCCTCCAG 1149
Qy 373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
Db 1150 CCCCACCCCGGAGTCACTCTGTACAGTACCCCTCAGACCTCCCGTACATGGGCTCC 1209
Qy 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
Db 1210 TATCACGCCACCCACAGAGATGAATTTGTGGCGCCCGCCACCTCCAGCCCTCCCGCTG 1269
Qy 413 ThrSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIle 432
Db 1270 ACATCTCCAGTTTGTGTGGCCCAACCCATAGTGGATTCCACCACTGGGGGTATA 1329
Qy 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1330 TACCCCAACTAGGCTCCCGCCACGACCATATGCTTCTCATCTGGGCACTTACTAC 1386

RESULT 5
US-10-205-823-98
; Sequence 98, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: W01-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
```

```
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 3166
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-98

Alignment Scores:
Pred. No.: 1,086-238 Length: 3166
Score: 2182.00 Matches: 407
Percent Similarity: 91.72% Conservative: 14
Best Local Similarity: 88.67% Mismatches: 24
Query Match: 89.28% Indels: 14
DB: Gaps: 4

US-09-902-772-2 (1-451) x US-10-205-823-98 (1-3166)
Qy 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
Db 281 GCAGCTCATATCAGGAAGCCCTTATCAGTTGTGAGTGAGGACCATGCTGTTTTCAGTGT 340
Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41
Db 341 GCGTACGGAAGCCACACCTGGCTAAGACAGAGATGACCGGTCTCTCCAGCGGACTAT 400
Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61
Db 401 GGACAGATTTCCAGATGAGCCCGCTGCTCCTCAGCAGGATGGCTGTCTCAACCCCA 460
Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db 461 GCGAGGGTCAACATCAAAATGGAATGTAACCTTAGCCAGGTGAATGGCTCAAGGAACCT 520
Qy 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly 101
Db 521 CTTGATGAATGCAATGTGGCCAAAGCGGGGAAGATGTGGGCGAGCCAGACACCGTTGG 580
Qy 102 MetAsnTyrGlySerTyrMetGluLysHisIleProProProAsnMetThrThrAsn 121
Db 581 ATGAACCTACGCGAGCTACATGAGGAGAGGACATGCCCCCAACATCACCACCAAC 640
Qy 122 GluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141
Db 641 GAGCGCAGAGTTATGCTGCCAGCAGATCTTACGCTATGGAGTACAGACCATGTCGGCAG 700
Qy 142 TrpLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
Db 701 TGGCTGGAGTGGCGGTGAAGAATAATGGCTTCCAGACGTCACATCTTGTATTCCAG 760
Qy 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
Db 761 AACATCGATGGGAAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCCC 820
Qy 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
Db 821 AGCTACAACCGCAGCATCTTCTCTCATCTCCTACCTACCTCAGAGAGACTCTCTTCCA 880
Qy 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
Db 881 CATTTG-----ACTTCA-----GATGATGTTGATAAAGCCTTCAAAACTCTCCA 925
Qy 222 -----AspLeuProTyrGluGlnAlaArgSerAlaTTP 233
Db 926 CGGTTAATGATGCTAGAAACACAGATTATCATATGAGCCCCCAGGAGATCAGCGCTGG 985
```

Score:	2182.00	Matches:	407
Percent Similarity:	91.72%	Conservative:	14
Best Local Similarity:	88.67%	Mismatches:	24
Query Match:	89.28%	Indels:	14
DB:	15	Gaps:	4
US-09-902-772-2 (1-451) x US-10-021-660-26 (1-3166)			
Qy 2	AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21		
Db	281 GCAGCTCATATCAAGGAAGCGTCTATCAGTTGTGAGTGAGCAGCAGTCTGTTGTGAGTGT 340		
Qy 22	AlaTyrcIysSerProHisLeuAlaLysThrGluuMetThrAlaSerSerSerSerGluTy 41		
Db	341 GCCTACGGAACGCCACACCTGGCTAGACAGAGATGACCGCGTCTCTCCAGCATAT 400		
Qy 42	GlycInThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProPro 61		
Db	401 GGACAGACTTCCAAGATGAGCCACGCGTCCCTCAGCAGGATTTGGTGTCTCAACCCCCA 460		
Qy 62	AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81		
Db	461 GCCAGGTCACCATCAAAATGGAAATGTAAACCTTAGCCAGGTGAATGGCTCAAGGAAC 520		
Qy 82	ProAspAaspCysSerValAlaLysGlyGlyMetValSerSerSerSerAspAsnValGly 101		
Db	521 CCGATGAATGAGTGTGGCCAAAGCGGGAAGATGGTGGCGACCCAGACACCGTTGG 580		
Qy 102	MetAsnTyrcIysSerTyrcMetGluGluLysHisIleProProAsnMetThrThrAsn 121		
Db	581 ATGAACATACGGCAGCTACATGSGAGGAAGACATACCCACCCCAACATGACACGAC 640		
Qy 122	GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141		
Db	641 GAGCGCAGATTATCGTCCAGCAGATCTTACCGTATGGAGTACAGACCATGTCGGCAG 700		
Qy 142	TrpLeuGluTrpAlaValLysGluTyrcGlyLeuProAspValAspIleLeuLeuPheGln 161		
Db	701 TGGCTGAGTGGCGGCTGAAAGAAATATGGCGTCTCCAGACGCTCAACATCTGTTATTCC 760		
Qy 162	AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181		
Db	761 AACATCATGGGAAGGAACGTGCGAGATGACCAAGGACGACCTTCCAGAGGCTCACCC 820		
Qy 182	SerTyrcAsnAlaAspIleLeuLeuSerHisLeuHisTyrcLeuArgGluArgGlyAlaThr 201		
Db	821 AGCTACACGCCGACATCTCTCTCACAATCTCCACTACCTCAGAGAGACTCCTCTCCA 880		
Qy 202	PheIlePheProAsnThrSerValTyrcProGluAlaThrGlnArgIleThrThrArgPro 221		
Db	881 CATTTG-----ACTTCA-----GATGATGTTGATAAAGCCTTACAAAACCTCTCCA 925		
Qy 222	-----AspLeuProTyrcGluGlnAlaArgArgSerAlaTrp 233		
Db	926 CGGTTAATGATGCTAGAAACACAGATTTACCATAATGAGCCCCCAGGAGATCAGCCTGG 985		
Qy 234	ThrSerHisSerHisProThr-----GlnSerLysAlaAlaThrGlnProSerSerThrVal 252		
Db	986 ACCGGTCACGGCCACCCCGCCCGAGTGTGAAAGCTGCTCAACCATCTCCTCTCCACAGT 1045		
Qy 253	ProLysThrGluAspGlnArgProGlnLeuAspProTyrcGlnIleLeuGlyProThrSer 272		
Db	1046 CCCAAACTGAGACACAGCGTCTCTCAGTTAGATTCCTTATCAGATTCCTTGGACCAACAGT 1105		
Qy 273	SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 292		
Db	1106 AGCGCGCTTCAATCCAGCAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGCTC 1165		
Qy 293	LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys 312		
Db	1166 CTGTGGACAGCTTCCAATCTCAGCTGCATCACCCTGGGAAGGCCCAACGCGGAGTTCAAG 1225		
Qy 313	MetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMet 332		

```
Db 1226 ATGACGGATCCCGACGAGGTGGCCGCGTGGGAGAGCGGAAGCAAAACCCACATG 1285
QY 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLys 352
Db 1286 AACTAGATAAGCTCAGCGCGCCCTCCGTTACTACTATGACAAGAACATCATGACCAAG 1345
QY 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
Db 1346 GTCCATGGGAAGCGCTACGCTACAAAGTTCGACTTCCACGGGATGCCCCAGGCCCTCCAG 1405
QY 373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
Db 1406 CCCCACCCCGCGAGTCATCTGTGTAAGTACCCCTCAGACCTCCCGTACATGGGCTCC 1465
QY 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
Db 1466 TATCAGGCCACCAACAGAGATGAATTTGTGGGCCCCACCTCCAGCCCTCCCGTG 1525
QY 413 ThrSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThrGlyIle 432
Db 1526 ACACTTCCAGTTTCTTGTGCCCCAACCCCATACTGGAATTCACCAACTGGGGTATA 1585
QY 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1586 TACCCCAACTAGGTCCCAACGACCATATGCTTCTCATCTGGGCACCTTACTAC 1642
```

RESULT 7

```
US-10-087-192-1124
; Sequence 1124, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION: David W.
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1124
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-1124
```

Alignment Scores:

```
Pred. No.: 5,74e-225 Length: 1297
Score: 2057.00 Matches: 384
Percent Similarity: 83.69% Conservative: 11
Best Local Similarity: 81.36% Mismatches: 9
Query Match: 84.17% Indels: 68
DB: 13 Gaps: 2
```

US-09-902-772-2 (1-451) x US-10-087-192-1124 (1-1297)

```
QY 7 GluAlaLeuSerValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26
Db 2 GAGGCCITGTCAGTTGTGAGCGAGGACCAAGTCACATTTTCAGTGTGCTACGGAACGCCA 61
QY 27 HisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyrGlyGlnThrSerLys 46
Db 62 CACCTGGCTAAGACAGAGATGACCGCATCTCTTCCAGTGAATGCGCCAGATCCCAAG 121
QY 47 MetSerProArgValProGlnGlnAspTripleSerGlnProProAlaArgValThrIle 66
Db 122 ATGAGTCCACAGTCCCTCAGCAGGACTGGCTGTCTCAAGCCCGCAGCGGTCAACATC 181
```

```
QY 67 LysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAspCysSer 86
Db 182 AAGATGGAGTGCAACCTAGTCAGTGAATGGTTCAGGAACCTCACCTGATGAGTGAGT 241
QY 87 ValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySer 106
Db 242 GTGAACAAAGGTGGAGATGTTGGGAGCCCGGATGCTGTGGGAGATGAGTACGCGAGC 301
QY 107 TyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgValIle 126
Db 302 TACATGGAGGAGAGCATGTGCGCCCTCCCAATATGACCACAAATGAGCGCAGAGTGATC 361
QY 127 ValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAla 146
Db 362 GTCCCTTCAGATCCCTACTCTGTGGAGCAGACCATGTCCGACAGTGGCTGGAGTGGCG 421
QY 147 VallysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys 166
Db 422 GTGAAGAAATATGGCTCTCGATGTGGAGCTTACTATTTCAGAAATATGATGGGAAG 481
QY 167 GluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAsp 186
Db 482 GAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGGCTCAGCGCAGTACAAATGCCGAC 541
QY 187 IleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
Db 542 ATTCTTCTCCATCTCCACTACTCCAGAGACTCCCTTCCACATCTGACTTCCGAT 601
QY 199 -----Gly 199
Db 602 GACGTTGATAAGGCTTTACAAAACCTCTCCACGGTTAATGTCATGCCAGAAACACAGGGGT 661
QY 200 AlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThr 219
Db 662 GCAGCTTTATTTTCCCAATACCTTCAATATCCGAGCTAGCAAGAAATATCAACT 721
QY 220 ArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisPro 239
Db 722 AGGCCA ----- 727
QY 240 ThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArg 259
Db 727 ----- 727
QY 260 ProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGly 279
Db 728 -----GATCCTTACCATCTCTGGACCGACCATGAGCCGCTTGTCTAATCCAGGT 778
QY 280 SerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuSerAspSerSerAsnSer 299
Db 779 AGTGCCAGATCCAGCTGTGGCAGTCTCTGCTCGAACTCCTGTGACAGAGCTCCAACTCC 838
QY 300 AsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluVal 319
Db 839 AACTGCATCACCCTGGGAAGGCCAACCGGGAGTTCAAGATGACAGACCCCGAGAGGTG 898
QY 320 AlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArg 339
Db 899 GCTCGGCGCTGGGGGAGAGGAAGCAAGCCCAACATGAATATGCAAGCTCAGCGCGC 958
QY 340 AlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAla 359
Db 959 GCCCTCCCGCTACTACTACGACAAAACATCATGACCAAGGTGCACGGAAGCGCTACGCC 1018
QY 360 TyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSer 379
Db 1019 TACAAGTTTGACTTCCACGGGATTTGCCAGGCCCTTGACGCCCACTCTGAGTCGTCC 1078
QY 380 MetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLys 399
Db 1079 CTGTAAAGTACCCCTCCGACCTGCATACATGGGCTCCTATCAGCCCACTCCCGCAGAG 1138
QY 400 MetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAla 419
```

Db 1139 ATGAACCTTTGTCTGCCACCTCCCGCTCTCCAGTCACATCTTCCAGTTCTTTGCT 1198
Qy 420 AlaProAsnProTyrTrpAsnSerProThrGlyGlyLeuValProAsnThrArgLeuPro 439
Db 1199 TCCCGAAGCACTACTGGAATTACCGACTGGGGGATCTACCCGAACACTAGGCTCCCA 1258
Qy 440 AlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1259 GCCAGGCATATGCCCTCTACCTGGGCACCTACTAC 1294

RESULT 8

US-10-052-482-198
; Sequence 198, Application US/10052482
; Publication No. US2004007264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 198
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-198

Alignment Scores:
Pred. No.: 1.87e-175 Length: 1359
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
Gaps: 7
DB: 12

US-09-902-772-2 (1-451) x US-10-052-482-198 (1-1359)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 1 ATGACGGGACTATTAGGAGGCTCTCGTGTGTGAGCGACGACGACGACGACGACGAC 60
Qy 21 CysAlaTyrGly----SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 61 TCAGCGTACGGAGCGGCGGCGCCATCTCCCAAGCGCGACATGACTGCTCGGGGAGTCCT 120
Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnAspTrpLeuSerGln 59
Db 121 GACTACGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db 181 ---CCAGTGAGGCTCAAGTCAAGCGGAGTAT-----GACCACATGAATGATCCAGG 231
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99
Db 232 GAGTCTCCGGTGGACTGAGCGGTATGCAATGACGAGGAGTGTGGCGGAGCGGAGTCC 291
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGlyHis----IleProProProAsnMet 118
Db 292 AACCCCATGAACATAACAGCTATATGACGAGAGAATGGCCCCCTCTCCCAACATG 351
Qy 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 352 ACCCAACGAGAGAGAGTATGCTGCTCCCGACACCCCACTGTGACACAGGAGCAT 411
Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 155

Db 412 GTGAGGCAATGGCTGGAGTGGGCGCATTAAGGAGTACAGCTTGATGGAGATCGACACATCC 471
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 472 TTTTTCAGACATGGATGGCAGAACTGTGTAATGAACAGAGGAGGACTTCTCCCGC 531
Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisIleuHisTyrLeuArgGluArg 198
Db 532 GCCACCACTCTTACACACGGAAGTGTGTGTCACACCTCAGTTACCTTACCTCAGGAA-- 588
Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db 589 ---AGTTCACTGCTGGCCTATAATACAACTCCACACCGACCAATCTCCAGATTGAT 645
Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHis 238
Db 646 GTCAAGAAGACCCCTTCTTATGACTCAGTCAGAGAGAGGCTTGGGCAATAAATGATGAT 705
Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
Db 706 TCTGGCCTCAACAAAGTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 765
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 766 CAACGGCCCGACGACGATCCGTATCAGATCTCTGGGCGGCGACGACGACGACGACGAC 825
Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerSerSer 297
Db 826 CTTGAGAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGGAGCTGCTCTCCGACAGCGCC 885
Qy 298 AsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 886 AACGCCAGCTGTATCACCCTGGAGGGGACCAACCGGGAGTTCAAAATGACGAGCCCCGAT 945
Qy 318 GluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 946 GAGGTGGCGGCGGCTGGGGGAGCGGAAAGCAAGCCCAACATGATATACGACAAAGCTG 1005
Qy 338 SerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1006 AGCCGGGCGCTCCGTTATTACTATGATATAAATGATGACCAAGTGCAGCGCAAAAGA 1065
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1066 TATGTTTCAAAATTTGACTTCCCGCATTCGCCAGGCTCTGCAGCCACATCCGACCGAG 1125
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1126 TCGTCCATGTACAGTACCTCTTACATCTCTTACATGCTCTCTACCATGCCACACAG 1185
Qy 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417
Db 1186 CAGAAAGTGAACTTTGCTCCCTCCCATCCATCCCTCCATGCTGCTCTCTCTCCAGTTTC 1245
Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db 1246 TTTGGAGCGCATCACATAGTACCTCCCGCGGGGGAATCTACCCCAACCCCAAC 1305
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1306 GTCCCGCGGCATCTTAACACCCACGCTGCTTTCACACTTAGCGAGCTACTAC 1356

RESULT 9

US-10-052-482-197
; Sequence 197, Application US/10052482
; Publication No. US2004007264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482

```

; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 197
; LENGTH: 2957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-197

Alignment Scores:
Pred. No.: 5,92e-175 Length: 2957
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
DB: 12 Gaps: 7

US-09-902-772-2 (1-451) x US-10-052-482-197 (1-2957)
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeupheGlu 20
DB 173 ATGGACGGGACTATTAGGAGGCTCTCGGTGGTGAGCGACGACGACGACGACGACGAC 232
QY 21 CysAlaTyrGly----SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
DB 233 TCACCGGTACGGGAGCGGCGCCACTCTCCCAAGGCGGACATGACTGCTCGGGAGTCT 292
QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTTPLeuSerGln 59
DB 293 GACTACGGGAGCGGCGCCAGATCAACCCCTCCACACGACGAGGAGTGGATCAATCAG 352
QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProGlnValAsnGlnSerArg 79
DB 353 ---CCAGTACGGGTCAAGCTCAAGCGGGAGTAT-----GACCAACATGAATGATCCAGG 403
QY 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
DB 404 GAGTCTCCGGTGGAGTGGAGGCTTAGCAATGACGAGGAGGCTGGTGGCGGGGAGTCC 463
QY 100 ValGlyMetAsnTyrGlySerTyrMetGluLysHis----IleProProAsnMet 118
DB 464 AACCCCATGAACACTCAACAGCTATATGACGAGAGGAAGTGGCCCTCTCTCCCAACATG 523
QY 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTTPSerThrAspHis 138
DB 524 ACCACCAACGAGAGGAGTATCGTCTCCCGACAGACCCACCTGTGGACACAGGAGCAT 593
QY 139 ValArgGlnTTPLeuGluTTPAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
DB 594 GTGAGGCAATGCTGGAGTGGCCATAAAGGAGTACAGCTTGTATGGAGATCGACACATCC 643
QY 159 LeupheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
DB 644 TTTTTCAGAAACATGATGGGAGGAACTGTGTAATGAACAGAGGAGGACTTCTCCCGC 703
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
DB 704 GCCACACCTCTCAACACGGAAGTGTGTGTGACACACCTCAGTACCTACCTCAGGAA-- 760
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
DB 761 ---AGTTCACTGCTGGCTCTATAATACACACCTCCACCGGACCAATCTCTCAGATTGAGT 817
QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTTPThrSerHisSerHis 238
DB 818 GTCAAGAGAGCCCTCTTATGACTCAGTCAAGAGGAGGCTTGGGCAATACATGAT 877
QY 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257

```

```

DB 878 TCTGCCTCAACAAAGTCTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 937
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaSer 277
DB 938 CAACGGCCCGGACGATCCGATCAGATCTCTGGGCGGACGAGCAGTCTGCTACCGAAC 997
QY 278 ProGlySerGlyGlnIleGlnLeuTTPGlnPheLeuLeuGluLeuLeuSerAspSerSer 297
DB 998 CTTGGAAGCGGGCAGATCCAGCTGTGGCAATCTCTCTGGAGTCTCTCCGACAGCGCC 1057
QY 298 AsnSerAsnCysIleThrTTPGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
DB 1058 AACGGCAGCTGTATCACCTGGGAGGGGACCAACCGGGAGTTTCAAAATACGAGCCCGAT 1117
QY 318 GluValAlaArgArgTTPGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
DB 1118 GAGGTGGCGGAGCGCTGGGGCGGCGGAAAGAACCCCAACATGAATACGACAGCTG 1177
QY 338 SerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
DB 1178 AGCCGGGCTCTCTCTTATTACTATGATAAAACATATGACCAAGTGCACGGCAAAAGA 1237
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
DB 1238 TATGCTTCAAAATTTGACTTCCAGCGCATTTGCCAGGCTCTGCACCCACATCCGACCGAG 1297
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
DB 1298 TCGTCCATGTACAACTACCTTCTGACATCTCTACATGCTCTCTACCATGCCACCAG 1357
QY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
DB 1358 CAGAAAGGTGAATTTGTCTCTCCATCCATCCATCCATCCATCCATCCATCCATCCAT 1417
QY 418 PheAlaAlaProAsnProTyrTTPAsnSerProThrGlyGlyIleTyrProAsn----- 435
DB 1418 TTTGAGCGGATCACTACATCTGAGCTCCCGACCGGGGGAATCTACCCCAACCCCAAC 1477
QY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyTyrTyrTyr 451
DB 1478 GTCCCGCGCATCTCAACACCCCGGCTGCTTCCACACTTAGGCGAGCTACTAC 1528

RESULT 10
US-10-342-887-710
; Sequence 710, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 710
; LENGTH: 2957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-710
Alignment Scores:

```

```
Pred. No.: 5,93e-175 Length: 2957
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
DB: 13 Gaps: 7

US-09-902-772-2 (1-451) x US-10-342-887-710 (1-2957)

Qy 1 MetAlaSerThrIleuValSerValValSerGluAspGlnSerLeuPheGlu 20
Db 173 ATGGACGGGACTATTAAAGAGGCTCTGCGGTGGTGAGCAGCAGGACCGCTCTTTGAC 232
Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 233 TCACGGTACGGAGCGGAGCCCATCTCCCAAGCCGACATGACTGCTCGGGAGTCT 292
Qy 40 GluTyrGlnThrSerLysMetSerProArgValProGlnInAspThrLeuSerGln 59
Db 293 GACTACGGGAGCGGAGCCCAAGATCAACCCCTCCCAAGCAGGAGGTGGATCAATCAG 352
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db 353 ---CCAGTGGGGTCAACGTCACGGGAGTAT-----GACCACATGATGGATCCAGG 403
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99
Db 404 GAGTCTCGGTGGAGCTGACGGCTTAGCAAAATGACCAAGCTGGTGGCGGAGGCGAGTCC 463
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---11eProProAsnMet 118
Db 464 AACCCTGAACTACACAGTATATGACGAGGAGAAATGGCCCCCTCTCCACATG 523
Qy 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 524 ACCACCAAGAGAGAGAGTATCGTCCCGCAGACCCACACTGTGGACACAGGAGCAT 583
Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
Db 584 GTGAGGCAATGGCTGGAGTGGCCATAAGAGAGTACAGCTTGATGGAGATCGACATCC 643
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 644 TTTTTCAGACATGATGGAGGAGACTGTGTAAATGAACAGGAGGACTTCTCCGCG 703
Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisIleuHisTyrLeuArgGluArg 198
Db 704 GCCACCACTCTACACACGGAAGTGTGTGTGTACACCTCAGTTACCTCAGGAA--- 760
Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db 761 ---AGTTCCTGCTGGCTATATACACCTCCACACCGGACCAATCTCAGATTGAGT 817
Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgGserAlaTrpThrSerHisSerHis 238
Db 818 GTCAAGAGACCCCTCTATGACTCAGTCAGTACAGAGAGGAGCTTGGGGCAATAACATGAT 877
Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerThrValProLysThrGluAsp 257
Db 878 TCTGGCTCAACAAAGCTCCCTCCCTGGAGGGGACCAACAGATCAGTAAATATACAGAG 937
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 938 CAACGGCCCCAGCCAGATCCGTATCAGATCTCGGGCCGACACAGCAGTGGCTAGCCAA 997
Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerSerSer 297
Db 998 CTTGAACGGGAGATCAGCTGGGCAATTCCTCTGGAGGCTCTCCGACAGGCGCC 1057
Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 1058 AACGCCAGCTGTATCACTGGGAGGGGACCAACGGGAGTTCAAAATACAGGACCCCGAT 1117

RESULT 11
US-10-172-118-710
; Sequence 710, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mac
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 710
; LENGTH: 2957
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 002017
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-710

Alignment Scores:
Pred. No.: 5,93e-175 Length: 2957
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
DB: 13 Gaps: 7

US-09-902-772-2 (1-451) x US-10-172-118-710 (1-2957)

Qy 1 MetAlaSerThrIleuValSerValValSerGluAspGlnSerLeuPheGlu 20
Db 173 ATGGACGGGACTATTAAAGAGGCTCTGCGGTGGTGAGCAGCAGGACCGCTCTTTGAC 232
Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
```

Db 233 TCAGCGTACGGAGCGGAGCCATCTCCCAAGGCGGACATGACTGCTCGGGAGTCT 292
Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGln 59
Db 293 GACTACGGGAGCCCAAGATCAACCCCTCCCAACAGGAGGAGTGAATCAATCAG 352
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db 353 --CCAGTGAGGGTCAACGTCAGCGGAGTAT-----GACCACATGAATGATCCAGG 403
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerAspAsn 99
Db 404 GAGTCTCCGGTGGACTGAGCGGTAGCAATAGCAGCAAGCTGTGGCGGAGCGAGTCC 463
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGlyHis---IleProProAsnMet 118
Db 464 AACCCCATGAATACACAGCTATATGACGAGAGAGATGGCCCCCTCTCCACATG 523
Qy 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHis 138
Db 524 ACCACCAACGAGAGAGATCATGCTCCCGCAGACCCACACTGTGGACACAGGAGCAT 583
Qy 139 ValArgGlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
Db 584 GTGAGGCAATGGCTGGAGTGGCCATPAGAGATGACAGCTGTGAGATCGACATCC 643
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 644 TTTTTCAGAACATGATGGTGGCAAGAACTGTGTAAATGAACAGGAGGACTTCTCCCG 703
Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisIleuHisTyrLeuArgGluArg 198
Db 704 GCCACACCTCTACACAGGAAGTCTGTGTACACCTCAGTACCTCAGCGAA--- 760
Qy 199 GlyAlaThrPheLeuPheProAsnThrSerValTyrProGlnAlaThrGlnArgIleThr 218
Db 761 --AGTTCACTGCTGGCTATATACAACTCCACACCGCAACATCTCAAGATGAGT 817
Qy 219 ThrArgProAspLeuProTyrGluAlaArgSerAlaThrThrSerHisSerHis 238
Db 818 GTCAAAGAACCCCTCTTATGACTAGTCAGTACAGAGAGAGCTGTGGGCAATAACATGAAT 877
Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerThrValProLysThrGluAsp 257
Db 878 TCTGGCTCAACAAAGTCTCCCTGGAGGGGACAAACGATCAGTAAGATAACAGAG 937
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleuGlyProThrSerSerArgLeuAlaAsn 277
Db 938 CAACGGCCCCAGCAGATCCGTATCAGATCTGGGCCCGCACAGCAGTCGCTAGCAAC 997
Qy 278 ProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuSerAspSerSer 297
Db 998 CTGAGAGCGGCGAGATCCAGCTGTGCAATCTCTCTGAGTGTCTCTCCGACAGGCC 1057
Qy 298 AsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 1058 AACGCCAGCTGATACCTGGAGGGACCAACCGGGAGTTCAAAATGACGAGCCCGCAT 1117
Qy 318 GluValAlaArgTyrGlyGluArgCysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 1118 GAGTGGCGCGCGCTGGGCGGCGGAAAGCAAGCCCAACATGATTAACGACAGCTG 1177
Qy 338 SerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1178 AGCCGGGCCCTCGTTATTATGATATAAAACATTATGACCAAGATGACGCGCAAGA 1237
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1238 TATCTTACAAATTTGACTTCCACGGCATTTGCCAGGCTTGACGCGCACATCCGACCGAG 1297
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1298 TCGTCCATGTACAGTACCCCTTCTGACATCTCTACATGCTTCTTACCATGCCACCG 1357

Qy 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhe 417
Db 1358 CAGAAGGTGAATTTGTCTCCCTCCATCTCTCCATGCTGTGCTCTCTCCAGCTTC 1417
Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db 1418 TTTGAGCGGATCAATACTATGACCTCCGCCACGGGGGAATCTACCCCAACCCCAAC 1477
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1478 GTCCCCCGCATCTCTAACACCCACGCTGCTTACACATTAGGAGCTACTAC 1528

RESULT 12

US-10-007-926A-295
; Sequence 295, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: PERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 295
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: friend leukemia virus integration 1 (FLI1)
; OTHER INFORMATION: Gene.
US-10-007-926A-295

Alignment Scores:

Pred. No.:	5,938-175	Length:	2957
Score:	1624.50	Matches:	306
Percent Similarity:	79.43%	Conservative:	57
Best Local Similarity:	66.96%	Mismatches:	83
Query Match:	66.47%	Indels:	11
DB:	15	Gaps:	7

US-09-902-772-2 (1-451) x US-10-007-926A-295 (1-2957)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 173 ATGGACGGGACTATTAAAGAGCTCTGCGTGGTGGAGCAGCAGCATCTCTTTTGC 232
Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 233 TCAGCGTACGGAGCGGCGAGCCCATCTCCCAAGGCGGACATGCTCGGGAGTCT 292
Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGln 59
Db 293 GACTACGGGAGCGGCGGAGGATCAACCCCTCCCAACAGGAGTGAATCAATCAG 352
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db 353 ---CCAGTACGGGTCAACGTCAAGCGGAGTAT-----GACCACATGAATGATCCAGG 403
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerAspAsn 99
Db 404 GAGTCTCCGGTGGACTGACAGCTTAGCAATAGCAGAAAGCTGGTGGCGGAGGAGTCC 463
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGlyHis---IleProProAsnMet 118

Db 589 ---AGTTCACTGCTGGGCTATAACACACCTCCCATACAGACCACTGCTCCACGACTGAAT 645
Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrrThrSerHisSerHis 238
Db 646 GTCAAGGAAGACCCCTTCTTATGACTCTGTCTCAGGAGAGGAGCATGGAACAATAATATCAAC 705
Qy 239 ProThrGlnSerLysAla---ThrGlnProSerSerSerThrValProLysThrGluAsp 257
Db 706 TCTGGCTCTCAACAAAGTCTCTCTCTGGAGGATCACAGACCATGGGCAAGAACACTGAG 765
Qy 258 GlnArgProGlnLeuAspProTyrGlnLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 766 CAGCGGCCCCAGCAGATCTCTTATCAGATCTCTGGGCCCAACACACAGCCGCTAGCAAC 825
Qy 278 ProGlySerGlyGlnLeuGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerSerSer 297
Db 826 CTTGGAGTGGGAGATCCAGCTGTGGCAGTTCTCTCTGGAACTACTGTCTCGACAGCGCC 885
Qy 298 AsnSerAsnCysIleThrTrpGluGlyThraSngGlyGluPheLysMetThrAspProAsp 317
Db 886 AACGCCAGCTGTATCACCTGGGAGGGACCAACGGGGAGTTTCAAAATGACGGACCTGAT 945
Qy 318 GluValAlaArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 946 GAGGTGGCCAGCGCTGGGAGAGCGGAGCAAGCCCAACATGATTTATGACAAGCTG 1005
Qy 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1006 AGCGGGGCCCTCCGATACTACTATGACAAACAAATATGACCAAGTGCATGGCAAAAGG 1065
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu 377
Db 1066 TATGCTTACAGTTTGAATTCATGGGATTCGCCAGGCCCTGCAGCCACATCCCAACAGAG 1125
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerTyrThrHisAlaHisPro 397
Db 1126 ACATCCATGACAGTATCCCTCTGATATCTCCATACATGCTCTCTACCATGCTCCATCAA 1185
Qy 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProAlaLeuProValThrSerSerPhe 417
Db 1186 CAGAAGGTGAATTTGCTCCGCTCACCCATCTCCATGCTCCATGCTGCTCCTCCAGCTTC 1245
Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db 1246 TTTGGAGCAGCATCAATACACTGGACCTCCGCCACTGCTGGGATCTATCCAAACCCAGT 1305
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1306 GTCCCCCGCATCTCTTAACACCCAGCTGCTTACACATTTAGCAGCTACTAC 1356

RESULT 14

US-10-052-482-194
; Sequence 194. Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 194
; LENGTH: 1729
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-052-482-194

Alignment Scores:

Pred. No.: 4,848-174 Length: 1729
Score: 1613.50 Matches: 305
Percent Similarity: 78.99% Conservative: 56
Best Local Similarity: 66.74% Mismatches: 85
Query Match: 66.02% Indels: 11
DB: 12 Gaps: 7
US-09-902-772-2 (1-451) x US-10-052-482-194 (1-1729)
Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 245 ATGCGCGGACTATTAGGAGGCTCTGCTGTGGTGGTGGAGTACGATCAGTCCCTTTTGTAT 304
Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 305 TCAGCATACGAGCGCGAGCCCATCTCCCAAGGAGCATATACACTGCTTCGGGGAGTCTCT 364
Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
Db 365 GACTACGGGAGCCCCACAAATACACCCCTGCCACCGCAGCAGGAGTGGATCAACACAG 424
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db 425 ---CCAGTGAGAGTCAATGTCAAGCGGAGTAT-----GACCACATGAATGCATCCAG 475
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99
Db 476 GAGTCTCCCGGTGAGCTGACAGTGTCAAGAAATGTAACAGCTGTGTGGCGGAGCGGAGCC 535
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProProAsnMet 118
Db 536 AACCCCATGACATATATATAGCTACATGATGATGAGAAGACGGCCCCCTCTCCCAACATG 595
Qy 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 596 ACCACCAAGCAAGCGGAGAGTCAATGTGCTGCAGACCCACACACTGTGGACAGCAGGAGCA 555
Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
Db 656 GTTCGACAGTGGCTGGAGTGTATAAGGAATACGGATTGATGGAGATTGACACTTCC 715
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 716 TTCTTCAGAACATGATGGCAGGAATTTGTAAATGAACAGAGGAGACTTCTCCGA 775
Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArg 198
Db 776 GCACCTCCGCTTACCAACAGAGAGTGTGTGTGCACCTCAGTTACCTCAGGGA--- 832
Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db 833 ---AGTTCACTGCTGGCTTATAACACACCTCCCATACAGACCACTCTCCAGACTGAAT 889
Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrrThrSerHisSerHis 238
Db 890 GTCAGGAAGACCTTCTTATGACTCTGTCCAGAGAGGAGCATGGACAATATATGAC 949
Qy 239 ProThrGlnSerLysAla---ThrGlnProSerSerSerThrValProLysThrGluAsp 257
Db 950 TCTGGCTCAACAAAGTCTCTCTCTGGAGGATCACAGACCATGGGCAAGAACACTGAG 1009
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgIleAlaAsn 277
Db 1010 CAGCGGCCCCAGCAGATCTTATCAGATCTCTGGGCCCAACAGCAGCGCCCTAGCAAC 1069
Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerSerSer 297
Db 1070 CTTGGAGTGGGAGATCCAGCTGTGGCAGTTTCTCTGGAATCTACTGTCCGACAGCGCC 1129
Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 1130 AACGCCAGCTGTATCACCTGGGAGGGGACCAACGGGGAGTTTCAAAATGACGAGCCCTGAT 1189

```

Qy 318 GluValAlaArgArgTrrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 1190 GAGTGGCCAGGCGCTGGGAGAGCGGAGCAAGCAAGCCCAACATGATATGACAAAGCTG 1249
Qy 338 SerArgAlaLeuArgTrrpTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1250 AGCGGGCCCTCCGATACTACTATGACAAACAAATATGACCAAGTGCATGCGCAAAAGG 1309
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1310 TATCCCTACAGTTTGACTTCCATGGCATTGCCAGGCCCTGCAGGCCACATCCACAGAG 1369
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1370 ACATCCATGTACAAAGTATCCCTCTGATATCTCTACATGCTTCTACCATGCCATCAA 1429
Qy 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417
Db 1430 CAGAGGTGAAGTCTTGGTCCCGTCTCACCCATCTCTCCATGCTGTACCTCTCCAGCTTC 1489
Qy 418 PheAlaAlaProAsnProTyrTrrpAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db 1490 TTTGGAGCAGCATCACAACTACTGGACCTGCCCACTGCTGGGATCTATCCAAACCCCAAGT 1549
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1550 GTCCCCGCCATCTTAACACCCAGCTGCTTCACACTTAGGCACTACTAC 1600

RESULT 15
US-10-087-192-1126
; Sequence 1126, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1126
; LENGTH: 212231
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1126

Alignment Scores:
Pred. No.: 9,05e-96 Length: 212231
Score: 956.50 Matches: 184
Percent Similarity: 79.75% Conservative: 9
Best Local Similarity: 76.03% Mismatches: 31
Query Match: 39.14% Indels: 18
Dbs: 13 Gaps: 3

US-09-902-772-2 (1-451) x US-10-087-192-1126 (1-212231)
Qy 227 GlnAlaArgArgSerAlaTrrpThrSerHisSerHisProThrGlnSerLysAlaThr-- 245
Db 201508 CAGGCAAGGCGAGTTCTCTGGAGAAGTGGGTGGCCATGGACGAGACGAGTCCGCAACCC 201567
Qy 246 -----GlnProSerSerThrValProLysThrGluAspG1 258
Db 201568 TGCAGCTAGGAGTGGACGAGCAGCCCTGGTGAAGGGAATGTGGCGGCACACCAATAGC 201627
Qy 258 nArgProGln-----LeuAspProTyrGlnIleLeuGlyPr 270

```

Search completed: May 28, 2004, 01:33:43
Job time : 799 secs

```

Db 201628 ACCTCCTTTCCGTCAATCAGAGAGCGCTCTTTCCCTCGGAACCTGTG---CTCAGCTGGAGT 201684
Qy 270 oThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrrpGlnPheLeuLe 290
Db 201685 TTTCTCTTCTCTCCTGACCTCATTAGGCAGTGGCCAGATCCAGCTTTGCGAGTTCTCTCT 201744
Qy 290 uGluLeuLeuSerAspSerSerAsnSerAsnCyalleThrTrrpGluGlyThrAsnGlyG1 310
Db 201745 GGAGCTCTCTGTGGACAGCTCCAACTCCAGTGTGATCACCTGGGAAGGACCAACGGGGA 201804
Qy 310 uPheLysMetThrAspProAspGluValAlaArgArgTrrpGlyGluArgLysSerLysPr 330
Db 201805 GTTCAAGATGACGGATCCCGACGAGTGGCCCGCGCTGGGGAGAGCGGAAGACAAACC 201864
Qy 330 oAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTrrpTyrTyrAspLysAsnIleMe 350
Db 201865 CAACATGAACACTACGATAGCTCAGCGCGCCCTCCGTTACTACTATGACAAAGAACATCAT 201924
Qy 350 tThrLysValHisGlyLysArgTrrpAlaTrrpLysPheAspPheHisGlyIleAlaGlnAl 370
Db 201925 GACCAAGGTCCATGGGAAGCGCTAGCGCTACAAGTTCGACTTCCACGGGATCGCCAGGC 201984
Qy 370 aLeuGlnProHisProGluSerSerMetTyrLysTrrpProSerAspLeuProTrrpMe 390
Db 201985 CCTCCAGCCCCACCCCGGAGTCTATCTCTGTACAAAGTACCCCTCAGACCTCCCGTACAT 202044
Qy 390 tSerSerTrrpHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLe 410
Db 202045 GGGCTCCTATACGGCCCAACCCACAGAAATGAACTTTGTGGCGCCCACTCCAGCCCT 202104
Qy 410 uProValThrSerSerSerPheAlaAlaProAsnProTrrpAsnSerProThrG1 430
Db 202105 CCCCCTGACATCTTCCAGTTTTTTTGTGTCGCCCAACCCATACCTGGAATTCACCAACTGG 202164
Qy 430 yGlyIleTrrpProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTy 450
Db 202165 GGGTATATACCCCAACACTAGGCTCCCAACAGCATATGCTTCTATCTCTGGGCACTTA 202224
Qy 450 rTyr 451
Db 202225 CTAC 202228

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 27, 2004, 20:54:02 ; Search time 3644 Seconds
(without alignments)
3695.896 Million cell updates/sec

Title: US-09-902-772-2
Perfect score: 2444
Sequence: 1 MASTIKKALSVSEDSQSLFE.....IYNTRLPAAHMSHLGTYT 451

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USP001/US0902772/runat_27052004_162841_2156/app_query.fasta_1.647
-DB=EST -OPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOPECL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0902772@cgn_1_3437_0runat_27052004_162841_2156 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORE=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2168.5	88.7	3064	11	AK078113	AK078113 Mus muscu
2	2157	88.3	1350	29	AY419454	AY419454 Homo sapi
3	2151	88.0	1866	11	AK050922	AK050922 Mus muscu
4	2126	87.0	1350	29	AY419456	AY419456 Mus muscu
5	1935	79.2	1350	29	AY419455	AY419455 Pan trogl
6	1613.5	66.0	1706	11	AK042111	AK042111 Mus muscu
7	1610.5	65.9	1347	29	AY418893	AY418893 Homo sapi
8	1602	65.5	3118	11	AK036655	AK036655 Mus muscu
9	1599.5	65.4	1347	29	AY418895	AY418895 Mus muscu
10	1428	58.4	1256	29	AY418894	AY418894 Pan trogl
11	1275	52.2	726	13	BU334592	BU334592 603500995
12	1123.5	46.0	709	14	CB269538	CB269538 1008445 H
13	1118.5	45.8	827	9	AJ456498	AJ456498 AJ456498
14	1055	43.2	2420	11	AK083900	AK083900 Mus muscu
15	1040	42.6	1845	12	EM456833	EM456833 AGENCOURT
16	1016.5	41.6	880	13	BQ233264	BQ233264 AGENCOURT
17	1002.5	41.0	1089	13	BQ212127	BQ212127 AGENCOURT
18	999	40.9	715	12	BG388025	BG388025 602412867
19	985	40.3	783	14	CD758623	CD758623 AGENCOURT
20	985	40.3	949	12	BI558012	BI558012 603240720
21	984	40.3	642	12	BI713036	BI713036 ie01d08.Y
22	984	40.3	668	10	B6660034	B6660034 B6660034
23	980	40.1	592	12	BM088063	BM088063 501436 MA
24	975.5	39.9	915	12	BG390291	BG390291 602416255
25	971.5	39.8	701	9	AU136709	AU136709 AU136709
26	970	39.7	629	10	AW948986	AW948986 QV4-FT000
27	966	39.5	1074	13	BQ953977	BQ953977 AGENCOURT
28	961	39.3	606	13	BU478556	BU478556 603471895
29	948	38.8	675	10	B6339043	B6339043 B6339043
30	947.5	38.8	634	12	BM489636	BM489636 pgmzh.BK0
31	941	38.5	616	14	CB579500	CB579500 AMGNNUC.N
32	938.5	38.4	976	13	BX328036	BX328036 BX328036
33	935.5	38.3	928	12	BG259459	BG259459 602378556
34	932	38.1	602	14	CB582161	CB582161 AMGNNUC.N
35	932	38.1	629	10	B577887	B577887 B577887
36	886	36.3	633	13	EX100094	EX100094 EX100094
37	872.5	35.7	526	14	CA405791	CA405791 1001940 H
38	853.5	34.9	521	13	BU440626	BU440626 604144050
39	846	34.6	802	14	CB952600	CB952600 AGENCOURT
40	840.5	34.4	671	12	BI558530	BI558530 603240526
41	830	34.0	1201	9	AL554108	AL554108 AL554108
42	828.5	33.9	666	12	BG390572	BG390572 602416344
43	822	33.6	553	12	BG258664	BG258664 602471144
44	821	33.6	851	13	BU108155	BU108155 603110416
45	816.5	33.4	524	9	AI794504	AI794504 fc45b11.Y

ALIGNMENTS

RESULT 1
AK078113
LOCUS
DEFINITION
AK078113 3064 bp mRNA linear HTC 18-SEP-2003
Mus musculus adult male medulla oblongata cDNA, RIKEN full-length
enriched library, clone:6330501C04 product:ERG mRNA, mouse homolog
of Human ets-related gene ERG, transcript variant 1, full insert
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AK078113
AK078113.1 GI:26346964
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
PUBMED 10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED 11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED 11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3064)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Numaya, K., Numazaki, R., Ohno, M., Oheato, N., Okazaki, Y., Saito, R., Saiboh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 3064
/organism="Mus musculus"
/mol_type="mRNA"

REFERENCE
JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

COMMENT

FEATURES
source

QY 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArg----- 198
 Db 601 CCGAGCTACAATCCGACATCTCTCTCATCTCCACTACCTCAGAGAGACTCCCTT 660
 QY 198 ----- 198
 Db 661 CCACATCTGACTTCCGATGAGTGTGATAGGCTTTACAAAACCTCTCCAGGTTATGCA 720
 QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213
 Db 721 GCCAGAAACACAGGGGGTGCAGCTTTTATTTTCCCAATACTTCAGTATATCCGAAGCT 780
 QY 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTyr 233
 Db 781 ACCCAAGAAATTCACACTAGGCCA----- 804
 QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerThrValPro 253
 Db 805 -----GCTGCTCAGCCATCTCCCTCTCAGTGGCC 834
 QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273
 Db 835 AAAAAGTCAAGACAGCGGCTCCTCAGTTAGATCCTTACCAGATCCTGGGACCCAGTAGC 894
 QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuLeuLeu 293
 Db 895 CGCTTTGCTAATCCAGGTAGTGGCCAGATCCAGCTGTGCAGTTCCTGCTCGAATCTCTG 954
 QY 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
 Db 955 TCAGACAGCTCCCAACTCCAACTCATCACCTGGGAAGGCACCAACCGGGAGTTCAGATG 1014
 QY 314 ThrAspProAspGluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsn 333
 Db 1015 ACAGACCCGACAGGTGGCTCGCGCTGGGGGAGAGGAAGCAAGCCCAACATGAAC 1074
 QY 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 353
 Db 1075 TATGACAAGCTCAGCGCGCCCTCCGCTACTACTACGACAAAAACATCATGACCAAGGTG 1134
 QY 354 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
 Db 1135 CACGGGAAGCGCTACCGCTACAAAGTTGACTTCCACGGGATTGCCAGGCGCTTCGAGCC 1194
 QY 374 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerTyr 393
 Db 1195 CACCTCTCGAGTCGTCCTGTCACAGTACCCCTCCGACCTGCCATACATGGGCTCCAT 1254
 QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413
 Db 1255 CACGCCACCCCCAGAGAAGATGAACTTTGTCTCTCCCAACCCCTCCCGCTCTCCAGTACA 1314
 QY 414 SerSerSerPheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyr 433
 Db 1315 TCTTCCAGTTCTTGTCTCCCGAACCCCATCTGGAAATTCACCGACTGGGCGCATCTAC 1374
 QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1375 CCGAACACTAGGCTCCAGCAGCCATATGCCCTCTCAGCTGGGCACTACTAC 1428
 RESULT 2
 AY419454 1350 bp DNA linear GSS 17-DEC-2003
 LOCUS Homo sapiens ERG gene, VIRTUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AY419454
 VERSION AY419454.1 GI:39775411
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1350)

AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civeillo, D.R., Lu, F., Murphy, B.,
 Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 1350)
 AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civeillo, D.R., Lu, F., Murphy, B.,
 Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 TITLE Direct Submission
 JOURNAL Submitted (15-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 Location/Qualifiers
 source 1..1350
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <1..>1350
 /gene="ERG"
 /locus_tag="HCM6896"
 gene
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,61e-176 Length: 1350
 Score: 2157.00 Matches: 400
 Percent Similarity: 91.09% Conservative: 9
 Best Local Similarity: 89.09% Mismatches: 36
 Query Match: 88.26% Indels: 4
 Db: 29 Gaps: 2
 US-09-902-772-2 (1-451) x AY419454 (1-1350)
 QY 7 GluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26
 Db 1 GAAGCTTATCAGTTGTGAGTGAGGACCACTCGTTGTTGAGTGTGCTACGGAACGCCA 60
 QY 27 HisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyrGlyGlnThrSerLys 46
 Db 61 CACCTGGCTAAGACAGAGATGACCGCGCTCTCTCCAGCGACTATGGACAGACTTCCAG 120
 QY 47 MetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaAlaTgValThrIle 66
 Db 121 ATGAGCCACACCGCTCCCTCAGCAGATTGGCTGTCTCAACCCCGACGCGGCTCACCATC 180
 QY 67 LysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAspCysSer 86
 Db 181 AAAATGGAAATGAACCCCTAGCCAGGTGAATGGCTCAAGGAACCTCTCTGTATGAATGCAGT 240
 QY 87 ValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySer 106
 Db 241 GTGGCCAAAGCGGGAGAGATGGTGGGACGCCACACCGTTGGATGAATACGCGGACG 300
 QY 107 TyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgValIle 126
 Db 301 TACATGGAGGAGAGACATGCCACCCCAACATGACCAAGCAAGCGCGCAGAGTTATC 360
 QY 127 ValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAla 146
 Db 361 GTGCCAGCAGATCCTCAGCTATGAGTACAGACCATGTGGCGCAGTGGCTGGAGTGGGCG 420
 QY 147 ValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys 166
 Db 421 GTGAAGAGATATGGCTTCCAGACGTCACATCTGTTATTCAGAACATCATGATGGGAAG 480
 QY 167 GluLeuCysLysMetThrLysAspPheGlnArgLeuThrProSerTyrAsnAlaAsp 186
 Db 481 GAACGTGCAAGATGACCAAGGACGACTTCCAGAGGCTTCCAGAGGCTCACCCCGCAGCTAC 540

QY 187 ILeuLeuSerHisLeuHisTyrLeuArgGlu-----ArgGlyAlaThrPheIle 203
 Db 541 ATCCCTCTCTCAACATCTCCACTACCTCGAGAGANNNNNNNNNNNNNNNNNNNNNNNN 600
 QY 204 PheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgProAspLeu 223
 Db 601 NNN 660
 QY 224 ProTyrGlnGlnAlaArgArgSerAlaTyrPheSerHisSerHisProThr---GlnSer 242
 Db 661 CCATATGAGCCCCCGAGAGATCAGCGCTGAGCCGCTCAGCCGACCCCGCCAGCTCG 720
 QY 243 LysAlaThrGlnProSerSerSerThrValProIleThrGlnAspGlnArgProGlnLeu 262
 Db 721 AAGAGTGTCAACATCTCTCTCAACATGCTGCCAAATGAAACCGAGCTCTCACTTA 780
 QY 263 AspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGln 282
 Db 781 GATCCTATCAAGATCTTGGACCAACAGTAGCCGCTTGCATATCCAGGACGTGGCCAG 840
 QY 283 ILeuGlnTyrGlnPheLeuLeuGlnLeuLeuSerAspSerSerSerSerSerSerValIle 302
 Db 841 ATCCGCTTGGCACTTCTCTCTGAGGCTCTGTCGACAGCTCCAACTCCAGCTGATC 900
 QY 303 ThrTyrGlnGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArg 322
 Db 901 ACCTGGAGAGGACCAACGGGAGTTCAAGATGACGATCCCGACGAGGTGGCCCGGCG 960
 QY 323 TrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArg 342
 Db 961 TGGGAGAGCGGAGAGCAAAACCAACATGACTAAGTAACTCAGCGCGGCGCTCCGT 1020
 QY 343 TyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe 362
 Db 1021 TACTACTATGACAGAAACATCATGACCAAGTCCATGGAAGCGCTAGCCTTACAGTTC 1080
 QY 363 AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlnSerSerMetTyrLys 382
 Db 1081 GACTTCCAGCGGATGGCCCGACGCTCCAGCCGCCCGGAGTCTCTCTTACAG 1140
 QY 383 TyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetAspPhe 402
 Db 1141 TACCCCTAGACCTCCGACATGGGCTCTCATATCGGCCACCCACCAAAATATAACTT 1200
 QY 403 ValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaAlaProAsn 422
 Db 1201 GTGGGCGCCCAACCTCCAGCCCTCCCGTACATTTTCAAGTTTTCGCGCCCAAC 1260
 QY 423 ProTyrTyrPheAsnSerProThrGlyIleTyrProAsnThrArgLeuProAlaHis 442
 Db 1261 CCATCTGTAATTCACCAACGTGGGGTATATACCCCAACCTAGGCTCCACAGGCAAT 1320
 QY 443 MetProSerHisLeuGlyThrTyrTyr 451
 Db 1321 ATGCTTCTCATCTGGGCACTTACTAC 1347

RESULT 3
 AK050922 1866 bp mRNA linear HTC 20-SEP-2003
 LOCUS AK050922.1 GI:26344597
 DEFINITION Mus musculus 9 days embryo whole body cDNA, Riken full-length
 enriched library, clone: D030036124 product: E18g mRNA, mouse homolog
 of human ets-related gene ERG, transcript variant 1, full insert
 sequence.
 ACCESSION AK050922
 VERSION AK050922.1 GI:26344597
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
 MEDLINE 20493374
 PUBMED 10493374
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komno, H., Akiyama, J., Nishi, K., Kikunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 JOURNAL sequencing pipeline with 384 multicapillary sequencer
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 MEDLINE 11076861
 REFERENCE 5
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 JOURNAL of 60,770 full-length cDNAs
 MEDLINE 11076861
 PUBMED 11076861
 REFERENCE 6
 AUTHORS (bases 1 to 1866)
 JOURNAL Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaki, A., Tanaka, Y.,
 Tanaka, T., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M., and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2001) Yoshitake Hayashizaki, The Institute of
 Biomedical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://pantom.gsc.riken.go.jp/
 URL: http://location/Qualifiers

FEATURES
 source
 1..1866
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM_DB:D030036124"
 /db_xref="MGI:2418825"
 /db_xref="taxon:10090"

CDS

```

/clone="D030036124"
/clone_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9 days embryo"
249..1637
/note="unnamed protein product; Brg mRNA, mouse homolog of
Human ets-related gene BRG, transcript variant 1
(DBJ|AB073078, evidence: BLASTN, 99%, match=1768)
putative"
/codon_start=1
/protein_id="BAC34461.1"
/db_xref="GI:26341598"
/translation="MIGVDPDAHIKELSVSEDSLEPCAVGTEPHLAKTEMTASS
SSDYGTSKMSPRVPOQDWLSQAPARVITKACNPSQVGSRRNSPDECSYKQKMGV
SPDVGMGSYGSMEKVPENMTNERRVVPADPTLMSIDHVRQWLEWAVEYGLL
DVIDLIPNDIGKELCKMTKDDPDLTPSVNADILSLHYLRTEPLPLHSDVDKA
LQNSPRLMHANTDLPYPRPARSAMTGHSHLTPOSKAQSPAPVETPEORPOLDY
QILGPTSRLLNPSGGIQLMQFLBELSSNSNCTTWSTNGEPFMTTPDVARW
GERSKPMWNTDKLSRALRYIDKNITKHKRYAKFDHGLAQMLQHPRESSLY
KYPEDLPYMGSYHAHPQKMFVSPHPALPVTSPPSPAPSPYVNSPTGIIYNTRLP
ASHMPSHLGITY"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 2,51e-175 Length: 1866
Score: 2151.00 Matches: 402
Percent Similarity: 91.07% Conservative: 16
Best Local Similarity: 87.58% Mismatches: 27
Query Match: 88.01% Indels: 14
DB: 11 Gaps: 4

```

US-09-902-772-2 (1-451) x AK050922 (1-1866)

```

QY 2 ALaSerThrIleYlgAlaLeuSeYValValSerGluAaspGlnSerLeuPheGluCys 21
DB 273 GCAGCTCATATTAAAGAGCGCTTGCAGTTGAGACGAGACCAAGTCACTATTGAGAGT 332
QY 22 ALaTYrGlySerProHisLeuAlaYThrGluMetThrAlaSerSerSerGluTYr 41
DB 333 GCCTAAGGAGACGCGCACCTGGCTAAGACAGAGAACCGCATCTCTCCAGGACAT 392
QY 42 GlnGlnThrSerIleMetSerProArgValProGlnGlnAaspTyrPleuSerGlnProPro 61
DB 393 GCGCCGACATCCAAATAGTAGTCCGAGTCCCTCGACGAGCACTGGCTGTCCAGGCCCA 452
QY 62 ALaArgValThrIleYMetGluCysAsnProAsnGlnValAsnGlySerIleAsnSer 81
DB 453 GCCAGGGTCACCATCAGATGAGTGAACCTTACTGAGTGAGTGAAGTTCAGGAACCA 512
QY 82 ProAspAspCysSerValAlaYlgGlyLysMetValSerSerSerAspAsnValGly 101
DB 513 CCTGATGAGTGAAGTGAACAAAGTGGAAAGTGGGAGGCCGAGATAGTGGGG 572
QY 102 MetAsnTYrGlySerTYrMetGlnGluLysHisIleProProProAsnMetThrThrasn 121
DB 573 ATGAGCTACGCGACGCTACATGAGAGAGCAATGGCCGCCCCCAATATGACCAAT 632
QY 122 GlnArgArgValIleValProAlaAspProThrLeuTyrPserThrAspHisValArgGln 141
DB 633 GAGCGAGAGTATGCTCCCTCGAATCTTCTCTGTGAGACAGACCAATGTCGACAG 692
QY 142 TrpLeuGlnTyrAlaValYlgGlyTYrGlyLeuProAspValAspIleLeuLeuPheGln 161
DB 693 TGGCTGAGTGGCGGTGAAGAAATATGGCTCTCGATGATGAGAGCTTACTATTATG 752
QY 162 AsnIleAspGlyLysGluLeuCYLysMetThrIleAspAspPheGlnArgLeuThrPro 181
DB 753 AATATGAGAGGAGAGAGCTGTGACAGACAAAGAGTCACTCCAGGCGCTCAGCGCG 812
QY 182 SerTYrAsnAlaAspIleLeuLeuSerHisLeuHisIleTYrLeuArgGlnArgGlyAlaThr 201
DB 813 AGCTTAAGTCCGACATTTCTTCTCTACATCTCCACTACCTCAGAGAGACTCCCTTCA 872

```

```

QY 202 PheIlePheProAsnThrSerValTYrProGlnAlaThrGlnArgIleThrThrArgPro 221
DB 873 CATCTG-----ACTTCC-----GATGACGTGATTAAGGCTTACAAATACTCTCCA 917
QY 222 -----AspLeuProTYrGlnGlnAlaArgArgSerIleATP 233
DB 918 CGGTTAATGCATGCCAGAAACACAGATTACTTATGAGCTTCCAGGAGATCAGCTTGG 977
QY 234 ThrSerHisSerHisProThr---GlnSerIleValThrGlnProSerSerSerThrVal 252
DB 978 ACCGGCCACAGCCACCTCATCCCTCAGTCCCAAGCTGCTCACCATCTCCTCTGCAAGT 1037
QY 253 ProLYrThrGlnAaspGlnArgProGlnLeuAaspProTYrGlnIleLeuGlyProThrSer 272
DB 1038 CCCMAAAGTGAAGACAGAGGCTCTCAGTTAGTCTTACACAGATCTGGAGCCGACAGT 1097
QY 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrGlnPheLeuLeuGlnLeu 292
DB 1098 AGCCGCTTGTCTATTCAGTGTAGTGGCCAGATCCAGCTGTGAGAGTCTGCTCGAATC 1157
QY 293 LeuSerAspSerSerAsnSerAsnCYLleThrTYrGlnGlyTYrAsnGlyGlnPheLys 312
DB 1158 CTGTCAAGACAGCTCCCACTCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 1217
QY 313 MetThrAspProAaspGlnValAlaArgArgTyrGlyGlyArgLysSerIleProAsnMet 332
DB 1218 ATGACAGACCCGAGCAGAGTGGCTCGGCTTGGGGAGAGAGAGACCAAGCCCAATG 1277
QY 333 AsnTYrAspLysLeuSerArgAlaLeuArgTYrTYrAspLysAsnIleMetThrLys 352
DB 1278 AACTATGACAGAGTCAAGCGCGCTCCGCTACTACTACGACAAACATCATGACCAAG 1337
QY 353 ValHisGlyLysArgTYrAlaTYrLysPheAspPheHisGlyIleHisGlnAlaLeuGln 372
DB 1338 GTGACAGAGGAGAGGCTACAGCTTCAAGTGTGACTTCACAGGAGTTCCAGGCCCTCGAG 1397
QY 373 ProHisProProGlnLeuSerMetTYrLysTYrProSerAspLeuProTYrMetSerSer 392
DB 1398 CCCACCTCTCTAGTGTCTTCCGTCAAGTACCTCCGACCTGCAATACAGGGCTCC 1457
QY 413 ThrSerSerSerPhePheAlaIleProAsnProTYrTYrAsnSerProThrGlyGlyIle 432
DB 1458 TATCAGGCCACCCCAAGAGTGAAGATTGTGTCCTCCACCTCCGCTCCCAATG 1517
QY 413 ThrSerSerSerPhePheAlaIleProAsnProTYrTYrAsnSerProThrGlyGlyIle 432
DB 1518 ACACTTCCAGTTCCTTGTCTTCCCGAACCCTACTGGAATTCAACGACTGGGGCATC 1577
QY 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyTYrTYr 451
DB 1578 TACCCGAACTAGTGGCTCCAGCCAGACCATATGCTCCCTCACTGAGCACTTACTAC 1634

```

RESULT 4
AY419456 1350 bp DNA linear GSS 17-DEC-2003
AX419456
LOCUS
DEFINITION Mus musculus ERG gene, VIRUAL TRANSCRIPT, partial sequence,
ACCESSION AY419456
VERSION AY419456.1 GI:39775413
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Taxonomy: Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1350)
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarirai,A.,
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
Interfering nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PubMed 14671302

Perriera, S., Wang, G., Zheng, X.H., White, T.J., Smirsky, J.J.,
Adams, M.D. and Cargill, M.

Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA

COMMENT
These sequences were made by sequencing genomic exons and ordering
them based on alignment.

FEATURES
source location/Qualifiers

1..1350

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

<1..>1350

/gene="ERG"

/locus_tag="HCM6896"

ORIGIN

Alignment Scores:

Pred. No.: 7,86e-157 Length: 1350
Score: 1935.00 Matches: 367
Percent Similarity: 83.74% Conservative: 9
Best Local Similarity: 81.74% Mismatches: 69
Query Match: 79.17% Indels: 4
D8: 29 Gaps: 2

US-09-902-772-2 (1-451) x AY19455 (1-1350)

QY 7 GUAAlaLeuSerValValSerGluApsGlnSerLeuPheGluCysAlaTyrGlySerPro 26
Db 1 GAAGCCTTATCATGATTGTGAGCGAGGACCACTGTTGAGTGTGCTACAGGAACGGCA 60
QY 27 HsLeuAlaValThrGluMetThrAlaSerSerSerSerGluTyrGlyGlnThrSerLys 46
Db 61 CACCTGGCTTAACAGAAATGACCGCGTCTCTCTCAAGCACTATGACAGCTTCCAG 120
QY 47 MetSerProAlaValProGlnGlnApsTyrPheLeuSerGlnProProAlaArgValThrIle 66
Db 121 ATGAGCCCAACGCGCTCTCAGAGGATGGCTGTCTCAACCCCGACGAGGCTCACATC 180
QY 67 LysMetGluCysAspProAlaGlnValAlaGlnGlySerArgAspSerProAlaApsCysSer 86
Db 181 AAAATGGAATGTAACTTACCTTACCGAGTGAACGGCTCAAGNNNNNNNNNNNNNNNNNNNN 240
QY 87 ValAlaIysGlyGlyLysMetValSerSerSerSerSerSerSerSerSerSerSerSer 106
Db 241 GTGGCCAAAGCGCGGAGATGTGGGCAAGCCAGACACCGTTGGATGAACACGCGAGC 300
QY 107 TyrMetGluGluLysHisGlnLeuProProProAlaMetThrThrAsnGluAlaArgValIle 126
Db 301 TACATGAGAGAGAGACATGCGNN 360
QY 127 ValProAlaApsProThrLeuTyrSerThrAspHisValArgGlnTyrPheGluTyrPala 146
Db 361 NNN 420
QY 147 ValIysGluTyrGlyLeuProAlaValAspIleLeuLeuPheGluAsnIleAspLysLys 166
Db 421 GTGAAGAATATGGCTTCCAGNNNTCAACATCANNNTTATTCAGAAACATCGATGGGAA 480
QY 167 GluLeuCysAlaMetThrLysAspApsPheGlnArgGluLeuThrProSerTyrAsnAlaAps 186
Db 481 NAACTGTGAGATGACCAAGGACGACTTCCAGAGGCTACCCCGAGTCAACAGCCGAGC 540
QY 187 IleLeuLeuSerHisLeuHisTyrLeuAlaGlu-----ArgGlyAlaThrPheIle 203
Db 541 ATCTCTCTCATCATCTCATCTCACTCAGAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
QY 204 PheProAlaThrSerValTyrProGluAlaThrGlnArgGleThrThrArgProApsLeu 223
Db 601 NNN 660
QY 224 ProTyrGluGlnAlaArgArgSerAlaTyrThrSerHisSerHisProThr--GlnSer 242

Db 661 CCATATGAGCCCCCGAGAGATCAGCTGAAACCGGTCAAGGCCACCCACGCCCCAGTGG 720
QY 243 LysAlaThrGlnProSerSerSerThrValProLysThrGluApsGlnArgProGlnLeu 262
Db 721 AAAGCTGCTCAACACATCTCTTCCACAGGCCCCAAACGAAACAGAGGTCTCTGATTA 780
QY 263 AspProTyrGlnIleLeuGlyProThrSerSerSerArgLeuAlaAsnProGlySerGlyGln 282
Db 781 GATCCTTATCATGATTTTGTGACCAACAGATAGCCGCTTGCAATTCAGGAGTGGCCAG 840
QY 283 IleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSerSerSerSerSerSerSer 302
Db 841 ATCCAGCTTGGAGATTCCTCTCCGAGAGCTCTGTGCGAAGAGCTCCCACTCCACTCATC 900
QY 303 ThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProApsGluValAlaArgArg 322
Db 901 ACCGGAGAGGACCAACGGGGAGTTCAAGATGACGATCCCAAGAGGTGGCCCGGCGC 960
QY 323 TrpGlyGluArgLysSerLysProAlaMetAsnTyrAspLysLeuSerArgAlaLeuArg 342
Db 961 TGGGAGAGGCGAAGAGCAAAACCAACATGAACTTACGATAGCTCAGCGCGCTCCGT 1020
QY 343 TyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe 362
Db 1021 TACTTACTATGACAAAGACATCATGACCAAGTCCATGGAGGCGTACGCTTACAAAGTTC 1080
QY 363 AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerMetTyrLys 382
Db 1081 GACTTCCACGGGATTCGCCAGGCGCTTCCAGCCCAACCCCGGAGATCTCTGTACAG 1140
QY 383 TyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetAsnPhe 402
Db 1141 TACCCCTCAACCTCCCGTATGAGGCTCTTACACAGCCCAACCAAGAGAACTT 1200
QY 403 ValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaIleProAsn 422
Db 1201 GTGGCGCCCAACCTCCAGCCCTCCCGTACATCTTCCAGTTTITTTGCTGCCCAAC 1260
QY 423 ProTyrTyrAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAlaIleHis 442
Db 1261 CCATACGTAATTCACCAACTGGGGTATATACCCCAATCTAGGCTCCCAACAGGCAT 1320
QY 443 MetProSerHisLeuGlyThrTyrTyr 451
Db 1321 ATGCCCTTCTCATCTGGGCACTTACTAC 1347

RESULT 6
AK042111
LOCUS
DEFINITION
Mus musculus 3 days neonate thymus cDNA, Riken full-length enriched library, clone: A63005911 product: Friend leukemia integration 1, full insert sequence.
AK042111
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL
MEDLINE
PUBMED
20499374
11042159

```

3  Shiba.K., Itoh.M., Aizawa.K., Nagaoka.S., Sasaki.N., Carninci.P.,
REFERENCE  Konno.H., Akiyama.J., Nishi.K., Kitsuai.T., Tashiro.H., Itoh.M.,
AUTHORS  Sumi.N., Iishi.Y., Nakamura.S., Hazama.M., Nishine.T., Harada.A.,
        Yamamoto.R., Matsumoto.H., Sakaguchi.S., Ikegami.T., Kashiwagi.K.,
        Fujiwaka.S., Inoue.K., Togawa.Y., Izawa.M., Ohara.E., Watahiki.M.,
        Oneda.Y., Ishikawa.T., Ozawa.K., Tanaka.T., Matsuda.S., Kawai.J.,
        Okazaki.Y., Murakatsu.M., Inoue.Y., Kira.A. and Hayashizaki.Y.
TITLE      RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL    sequencing pipeline with 384 multicapillary sequencer
MEDLINE     Genome Res. 10 (11), 1757-1771 (2000)
PUBMED      20530913
11076861

4  The RIKEN Genome Exploration Research Group Phase II Team and the
REFERENCE  FANTOM Consortium.
AUTHORS    Functional annotation of a full-length mouse cDNA collection
TITLE      Nature 409, 685-690 (2001)
JOURNAL
PUBMED

5  The FANTOM Consortium and the RIKEN Genome Exploration Research
REFERENCE  Group Phase I & II Team.
AUTHORS    Analysis of the mouse transcriptome based on functional annotation
TITLE      of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
PUBMED      12017061

6  (bases 1 to 1706)
REFERENCE  Adachi.J., Aizawa.K., Akimura.T., Arakawa.T., Bono.H., Carninci.P.,
AUTHORS    Fukuda.S., Furuno.M., Hanagaki.T., Hara.A., Hashizume.W.,
        Hayashida.K., Hayatsu.N., Hiramoto.K., Hiraoka.T., Hirozane.T.,
        Horii.F., Imotani.K., Ishii.Y., Itoh.M., Kagawa.T., Kasukawa.T.,
        Kato.H., Kawai.J., Kojima.Y., Kondo.S., Konno.H., Kouda.M.,
        Koya.S., Kurihara.C., Matsuyama.T., Miyazaki.A., Murata.M.,
        Nakamura.M., Nishi.K., Nomura.K., Numazaki.R., Ohno.M., Ohsato.N.,
        Okazaki.Y., Saito.R., Saitoh.H., Sakai.C., Sakai.K., Sakazume.N.,
        Sano.H., Sasaki.D., Shibata.K., Shinagawa.A., Shiraki.T.,
        Sogabe.Y., Tagami.M., Tagawa.A., Takahashi.F., Takaku-Akahira.S.,
        Takeda.Y., Tanaka.T., Tomaru.A., Toya.T., Yasunishi.A.,
        Muramatsu.M. and Hayashizaki.Y.
TITLE      Direct Submission
JOURNAL     Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
PUBMED      Physical and Chemical Research (RIKEN), Laboratory for Genome
        Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
        RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
        Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
        URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
        Fax: 81-45-503-9216)

COMMENT    cDNA library was prepared and sequenced in Mouse Genome
        Encyclopedia Project of Genome Exploration Research Group in Riken
        Genomic Sciences Center and Genome Science Laboratory in RIKEN.
        Division of Experimental Animal Research in Riken contributed to
        prepare mouse tissues.
        Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
        Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
        Trust/MRC building Addenbrookes Hospital Cambridge) whose
        assistance we gratefully acknowledge.
        Please visit our web site for further details.
        URL: http://genome.gsc.riken.go.jp/
        Location/Qualifiers
            1..1706
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="FANTOM,DB:A630059111"
                /db_xref="MGI:2407136"
                /db_xref="taxon:10090"
                /clone="A630059111"
                /tissue_type="thymus"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="3 days neonate"
            158..1516
                /note="unnamed protein product; Friend leukemia
                integration 1 (MGD|MGI:95554, GB|NM_008026, evidence:
                BLASTN, 99%, match=1671)

CDS

```

ORIGIN

```

Alignment Scores:      8.02e-129      Length:      1706
Pred. No.:            1613.50      Matches:      305
Score:                78.99%      Conservative:  56
Percent Similarity:   66.74%      Mismatches:   85
Best Local Similarity: 11          Indels:       11
Query Match:         11          Gaps:         7
DB:

US-09-902-772-2 (1-451) x AK042111 (1-1706)

QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
DB 158 ATGACGGGACTATTAAAGGAGCTCTCTGTGTGTGAGTGCAGCATCAGTCCCTTTTGTAT 217
QY 21 CysAlaTyrGly---SerProHieAlaLysThrGluMetThrAlaSerSerSer 39
DB 218 TCAGCATACGGGCGGCGGCCATCTCCCAAGGCAGATATGATGTGTCGGGAGTCTCT 277
QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
DB 278 GACTACGGGCGGCCCAAAATCAACCCCTGCGCACCGCAGCAGGAGTGTGATCAACCG 337
QY 60 ProProAlaArgValThrIleLysMetLysCysAsnProAsnGlnValAsnGlySerArg 79
DB 338 ---CCAGTGAGAGTCAATGTCAAGCGGAGTAT-----GACCACATGATGGATCCAGG 388
QY 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsn 99
DB 399 GAGTCTCGGTGGAGTCTGAGTGTGAGCAATGTAAACAGCTGTGTGGCGGAGGCGAAGCC 448
QY 100 ValGlyMetAsnTyrGlySerTyrMetClnGluLysHis---11eProProAsnMet 118
DB 449 AACCCCATGAATATAATAGTATGATGATGAGAGAACAGCCGCCCTCTCCCAACATG 508
QY 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
DB 509 ACCACCAACGACGGAGAGTCAATGTGCTGCTGACAGACCCACACACACAGGAGCAC 568
QY 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
DB 569 GTTCACAGTGGCTGGAGTGGGCTATATAAGGAATACGGATTCATGGAGATTCACACTTCC 628
QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
DB 629 TTCCTCCAGACATGATGGCAAGGAATGTGTAATATGAACACAGGAGGACTTCTCTCCGA 688
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArg 198
DB 689 GCCACCTCGCCTACACACAGAGAGTGTGTGTGCGCACCTCAGTTACCTCAGGGA--- 745
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
DB 746 ---AGTTCACTGCTGGCTTATACACCAACCTCCCAACAGACAGGACTCTCCAGACTGAT 802
QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHis 238
DB 803 GTCAGGAAGACCCCTTCITATGACTCTGTGACGAGGAGGAGCATGGAACAATAATATGAC 862

```

putative"

```

/codon_start=1
/protein_id="BAC31169.1"
/db_xref="GI:26334937"
/translations="MDGTIKKALSVVSDSLFSDSAYGAAHLPKADMTASGSPDYGQ
PHKINLPPOEWINQPVVVKREYDHNGSRESFVDCSVSKCNKLAVGGEANPMY
NSYMDKNGPPPNMTNERRVTPADPTLTQEHVROWLEWAIKEYGLMEIDTSPFO
NMGKELCVANKEDEFRATSAKTEVLLSHLSYRESSLLAYNTSHSTDOSSRLNKE
DPSIDVRRGANWNNNSGLPPLGGSTQMGKNTQEPQDPYQIOLGTSRRLANP
GSGQIQWQLLSLSDSANSCITWETNGEFGKATDDEVARWGERSKPMNVDK
LSRALYYDKNIMTVHGKRYAYKDFHGIAGLQHPHTETSMYKIPSDISWMPSVH
AHQQKVNFPVSHPSMPTVSSSPFGAASQYWTSTAGIYFNPSPVRPHENTHVPVSHLGS
YY"

```

QY	239	ProThrGlnSerLysAla---ThrGlnProSerSerSerThrValProLysThrGluAsp	257
Db	863	TCCTGGGCTCTCAACAAAGATCTCTCTCTGGAGGATCACAGACCATGGCAAGAACTGAG	922
QY	258	GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn	277
Db	923	CACGGCCCCCAGCCAGATCTCTTATAGATCCTGGGGCCAAACAGAGCCCTTAGCAAAAC	982
QY	278	ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSer	297
Db	983	CCTGGGAGTGGGCAGATCCAGCTGTGACGTTTCTCTGGAATCTACTGTCCGACAGCGCC	1042
QY	298	AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysValHisGlyLysArg	317
Db	1043	AACGCCAGCTGTATACCTCGGAGGGACCAACGGGGAGTTCAAAATGACGGACCCCTGAT	1102
QY	318	GluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu	337
Db	1103	GAGTGGCCAGGCGCTGGGGAGAGCGGAGCAAGCCCAACATGATTTATGACAAGCTG	1162
QY	338	SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg	357
Db	1163	AGCGGGGCCCTCCGATCTACTTATGACAAACAAATTTATGACCAAGTGCATGCCAAAGG	1222
QY	358	TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu	377
Db	1223	TATGCCCTACAGTTTGACTTCCATGGCATTGGCCAGGCCCTGCAGCCACATCCACAGAG	1282
QY	378	SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro	397
Db	1283	ACATCCATGTACAAATATCCCTCTGATATCTCTACATGCTCTCTACCATGCCATCAA	1342
QY	398	GlnLysMetAsnPheValAlaProHisAspProAlaLeuProValThrSerSerSerPhe	417
Db	1343	CAGAAGGTGAACATTGTCCCGTCTCACCCATCTCCATGCTGCATCCTCCTCCAGCTTC	1402
QY	418	PheAlaAlaProAsnProTyrTyrPAsnSerProThrGlyGlyIleTyrProAsn----	435
Db	1403	TTTGGAGCAGCATCACAAATCTGGACCTCCCCACTGCTGGATCTATCCAAACCCAGT	1462
QY	436	---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr	451
Db	1463	GTCCCCCGCATCTCTAACACCCAGTCCCTTCACATTAGCGAGCTACTACTAC	1513
RESULT 7			
LOCUS	AY418893	1347 bp DNA linear GSS 17-DEC-2003	
DEFINITION		Homo sapiens FLI1 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.	
ACCESSION	AY418893		
VERSION	AY418893.1	GI:39774853	
KEYWORDS	GSS.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS		Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.	
TITLE		Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios	
JOURNAL		Science 302 (5652), 1960-1963 (2003)	
PUBMED		14671302	
REFERENCE		2 (bases 1 to 1347)	
AUTHORS		Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.	
TITLE		Direct Submission	
JOURNAL		Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA	

Qy 262 LeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGly 281
 Db 766 CCAGATCCGATCAGATCTCTGGGCCCGACAGCAGTCGCTAGCCAACTGGAAGCGG 825
 Qy 282 GlnIleGlnLeuTyrGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCys 301
 Db 826 CAGATCCAGCTGTGGCAATTCCTCTGGAGTCTCTCCGACAGCCCAACGCCAGCTCT 885
 Qy 302 IleThrTyrGluGlyThrAsnGlyGluPheLeuSerMetThrAspProAspGluValAlaArg 321
 Db 886 ATCACCTGGAGGGAGCAACGGGGAGTTCAAAATGACGACCCCGATGAGTGGCCAGG 945
 Qy 322 ArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeu 341
 Db 946 CGTGGGGCGAGGGAACCAAGCCCAACATGATGACCAAGCTGAGCGGGCCCTC 1005
 Qy 342 ArgTyrTyrTyrAspLysAsnIleMetThrLysVaHisGlyLysArgTyrAlaTyrLys 361
 Db 1006 CGTATTACTATGATATAAAACATTATGACCAAGTGCACGGCAAAAGATATGTTACAA 1065
 Qy 362 PheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMetTyr 381
 Db 1066 TTTGACTTCCAGGCATGTCGCCAGGCTCTCCAGCCATCCGACGAGTGTCCAGTAC 1125
 Qy 382 LysTyrProSerAspLeuProTyrMetSerTyrHisAlaHisProGlnLysMetAsn 401
 Db 1126 AAGTACCCTTCTGACATCTCTACATGCTCTTCTACCATGCCACCAGCAGAGGTGAAC 1185
 Qy 402 PheValAlaProHisProAlaLeuProValThrSerSerPheAlaAlaPro 421
 Db 1186 TTTGTCCTCCCATCATCTCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1245
 Qy 422 AsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn-----ThrArgLeu 438
 Db 1246 TCACAACTACTGAGCTCCCCACAGGGGGAATCTACCCCAACCCCAACGTCGCCCAT 1305
 Qy 439 ProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1306 CCTAACACCCACGTCCTTCCACTTAGCGAGCTACTAC 1344

RESULT 8
 LOCUS AK036655 3118 bp mRNA linear HTC 19-SEP-2003
 DEFINITION Mus musculus adult male bone cDNA, RIKEN full-length enriched library, clone:9830146N17 product:Friend leukemia integration 1, full insert sequence.
 ACCESSION AK036655
 VERSION AK036655.1 GI:26331575
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
 2 20499374
 11042159
 3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

262 LeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGly 281
 766 CCAGATCCGATCAGATCTCTGGGCCCGACAGCAGTCGCTAGCCAACTGGAAGCGG 825
 282 GlnIleGlnLeuTyrGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCys 301
 826 CAGATCCAGCTGTGGCAATTCCTCTGGAGTCTCTCCGACAGCCCAACGCCAGCTCT 885
 302 IleThrTyrGluGlyThrAsnGlyGluPheLeuSerMetThrAspProAspGluValAlaArg 321
 886 ATCACCTGGAGGGAGCAACGGGGAGTTCAAAATGACGACCCCGATGAGTGGCCAGG 945
 322 ArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeu 341
 946 CGTGGGGCGAGGGAACCAAGCCCAACATGATGACCAAGCTGAGCGGGCCCTC 1005
 342 ArgTyrTyrTyrAspLysAsnIleMetThrLysVaHisGlyLysArgTyrAlaTyrLys 361
 1006 CGTATTACTATGATATAAAACATTATGACCAAGTGCACGGCAAAAGATATGTTACAA 1065
 362 PheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMetTyr 381
 1066 TTTGACTTCCAGGCATGTCGCCAGGCTCTCCAGCCATCCGACGAGTGTCCAGTAC 1125
 382 LysTyrProSerAspLeuProTyrMetSerTyrHisAlaHisProGlnLysMetAsn 401
 1126 AAGTACCCTTCTGACATCTCTACATGCTCTTCTACCATGCCACCAGCAGAGGTGAAC 1185
 402 PheValAlaProHisProAlaLeuProValThrSerSerPheAlaAlaPro 421
 1186 TTTGTCCTCCCATCATCTCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1245
 422 AsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn-----ThrArgLeu 438
 1246 TCACAACTACTGAGCTCCCCACAGGGGGAATCTACCCCAACCCCAACGTCGCCCAT 1305
 439 ProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 1306 CCTAACACCCACGTCCTTCCACTTAGCGAGCTACTAC 1344

AK036655 3118 bp mRNA linear HTC 19-SEP-2003
 Mus musculus adult male bone cDNA, RIKEN full-length enriched library, clone:9830146N17 product:Friend leukemia integration 1, full insert sequence.
 AK036655
 AK036655.1 GI:26331575
 HTC; CAP trapper.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
 2 20499374
 11042159
 3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection. Nature 409, 685-690 (2001)
 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
 6 (bases 1 to 3118)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayateu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. 3118
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM_DB:9830146N17"
 /db_xref="MGI:2401092"
 /db_xref="taxon:10090"
 /clone="9830146N17"
 /sex="male"
 /tissue_type="bone"
 /clone_id="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 192. 1547
 /note="unnamed protein product; Friend leukemia integration 1 (MGI:95554, GB|NM_008026, evidence: BLASTN, 99%, match=1671) putative"
 /codon_start=1
 /protein_id="BAC29518.1"
 /db_xref="GI:26331576"
 /translation="MDGTRKALSVVDDQSLFDSAYGAARHLPKADMTASGSPDYGQ
 PHKINPLPQQEWINQPVVNVKRYDNGNSRSPVDCVKNCLVGGGANPNYN
 SYMDKNGPPPNMTNRRVIVPADPLTQEHVRLQWLEWIGLMEIDTSFFQN
 MDGKELCKNKKEDFLRATSAVNTVLLSHLSYLRSSLLAYNTTTSHTDQSSRLNKED

FEATURES

source

CDS

PSYDVRGAWNNMNSGLKSLPLGGSTGKMTQEPQDPDQVILGPTSSRLANPG
SGOIQWQLELLSDSANASCIWETNGSPKOTDDEVARWSEKSKPNMNYDKL
SRALYVYKXNIMTKVKGKAYKFPDZHQIAQLOPHPTETSMYKVPDSISYMPHYA
HQQKVNFPVSPNPVTSISFFGAASQYWTISFTAGIYPNPSVPRHPNHPVSHLGSY
Y"

ORIGIN

Alignment Scores:

Pred. No.: 2,01e-127 Length: 3118
Score: 1602.00 Matches: 305
Percent Similarity: 78.77% Conservative: 55
Best Local Similarity: 66.74% Mismatches: 85
Query Match: 65.55% Indels: 12
DB: 11 Gaps: 8

US-09-902-772-2 (1-451) x AK036655 (1-3118)

QY 1 MetAlaSerThrIleLeuValSerGluAspGlnSerLeuPheGlu 20
Db 192 ATGGACGGGACTATTAAAGGAGCTCTGTCTGTGTGAGTACGATCAGTCCCTTTTGTAT 251
QY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 252 TCAGCATACGAGCGGCGAGCCCATCTCCCAAGGACAGATATGCTGCTTCGGGGAGTCCT 311
QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGln 59
Db 312 GACTACGGGCGACCCCAAAATCAACCCCTGCGCAGGAGGAGTGGATCAACAG 371
QY 60 ProProAlaArgValThrIleLeuMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db 372 ---CCAGTCAGAGTCATGTCAAGCGGAGTAT-----GACCACATGATGATCCAGG 422
QY 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99
Db 423 GAGTCTCCGGTGGAGCTGCAGTGTGTC---AAATGTAACAGTGTGTGGCGGAGGCGAAGCC 479
QY 100 ValGlyMetAsnTyrGlySerTyrMetCluGluLysHis----IleProProAsnMet 118
Db 480 AACCCCATGAACTAATAGTCTATGATGATGAGAGAACGCGCCCTCTCCCAACATG 539
QY 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHis 138
Db 540 ACCACCAACGAGGAGAGTCTGTGCTGCAGACCCACACCTGTGGACACAGGAGCAC 599
QY 139 ValArgGlnTyrLeuTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
Db 600 GTTCACAGTGGCTGGAGTGGCTATAAAGGAATACGGATTTGATGGAGATTGACACTTCC 659
QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 660 TTCITCCAGACATGATGGCAAGGAATGTGTAATGAACAAGGAGGACTTCTCCCGA 719
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
Db 720 GCCACCTCCGCTACACACAGAAGTGTGTGTGCGACCTCAGTACCTCAGGAA--- 776
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db 777 ---AGTTCAGTCTGGCTTAAACACACCTCCCATACAGACCTGCTCAGCTCAAT 833
QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTyrThrSerHisSerHis 238
Db 834 GTCAGGAAGACCCCTTCATGACTCTCTCAGGAGGAGGAGCATGGAACAATAATATGAAC 893
QY 239 ProThrGlnSerLysAla---ThrGlnProSerSerThrValProLysThrGluAsp 257
Db 894 TGTGCGCTCAACAAAGTCTCTCTCTGGAGGATCACAGACCATGGCGGCAAGAACACTGAG 953
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 954 CAGCGGCCCCCAGCCAGATCCCTTATCAGATCTCTGGGGCCAAACAGCAGCGCGCTAGCAAC 1013

QY 278 ProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSer 297
Db 1014 CTGGGAGTGGGACAGATCCAGCTGTGGCAGTTTCTCTGGAATCTACTGTCCGACAGCGCC 1073
QY 298 AsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 1074 AACGCCAGCTGTATCACCTGGGAGGGGACCAACGGGGAGTTCAAATGACGACCCCTGAT 1133
QY 318 GluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 1134 GAGTGGCCAGGCGCTGGGAGAGCGGAGAGCAAGCCCAACATGATATCACAGCTG 1193
QY 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1194 AGCCGGGCGCTCCGATACTACTATGACAAAAACATATGACCAAGTGCATGCAAAAGG 1253
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1254 TATGCTACAGTTTGACTTCCATGGCATTTGCCAGGCGCTGCAGCCACATCCACAGAG 1313
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1314 ACATCCATGTACAAGTATCCCTCTGTATCTCTCATATCTCTACATGCTCTCTACCATCAA 1373
QY 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417
Db 1374 CAGAAGTGAATTTGTCCCGTCTCACCATCTCCATCCCTGTCACCTCTCCAGCTTC 1433
QY 418 PheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db 1434 TTTGAGAGCATCAATACATCTGGACCTCCCTCCCTGCTGATCTATCAAAACCCAGT 1493
QY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1494 GTCCCCCGCATCTTAACACCCAGCTGCTTCCATCTAGGAGCTACTAC 1544

RESULT 9
LOCUS AY418895 1347 bp DNA linear GSS 17-DEC-2003
DEFINITION Mus musculus FLII gene, VIRUAL TRANSCRIPT, partial sequence,
Genomic survey sequence.
ACCESSION AY418895
VERSION AY418895.1 GI:39774855
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1347)
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
Gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1347)
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment
FEATURES
source 1..1347
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..->1347

/gene="FLJ11"		/locus_tag="HCM6709"	
ORIGIN			
Alignment Scores:			
Pred. No.:	9.1e-128	Length:	1347
Score:	1599.50	Matches:	301
Percent Similarity:	79.25%	Conservative:	58
Best Local Similarity:	66.45%	Mismatches:	83
Query Match:	65.45%	Indels:	11
DB:	29	Gaps:	7
US-09-902-772-2 (1-451) x AY418895 (1-1347)			
QY	5	IleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCysAlaTyrGly	24
DB	1	TTTCAGGAGGCTCTCTGTGTGAGTACAGATCAGTCCCTTTTGTATCAGCATACGGA	60
QY	25	---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyrGlyGln	43
DB	61	GCGGAGCCCATCTCCCAAGGAGATGATGACTGCTTCGGGGAGTCTGTACTACGGGCG	120
QY	44	ThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProProAlaArg	63
DB	121	CCCCACAAATCAACCCCTGCCACCGCAGCAGAGTGGATCAACACAG---CCAGTGAGA	177
QY	64	ValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAsp	83
DB	178	GTCAATGTCAAGCGGAGTAT-----GACCACATGATGGATCCAGGGAGTCTCCGGTG	231
QY	84	AspCysSerValAlaLysGlyGlyLysMetValSerSerSerSerSerAsnValGlyMetAsn	103
DB	232	GACTGCAGTGTCAAGAAATGTAACAGCTGTGTGGCGGAGCGCAAGCCCAACCCCATGAAC	291
QY	104	TyrGlySerTyrMetGluLysHis---IleProProProAsnMetThrThrAsnGlu	122
DB	292	TATTAATGCTACATGATGAGAGAACGGCCCCCTCTCTCCACATGACACCAACGAA	351
QY	123	ArgArgValIleValProAlaAspProThrLeuTyrPsrThrAspHisValArgGlnTyr	142
DB	352	CGGAGAGTCAATGTGCTGCAGACCCACACATGTGGACACAGGACAGCTTCCGACAGTGG	411
QY	143	LeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsn	162
DB	412	CTGGATGGGCTATAAAGAAATACGATTTGATGGAGATTGACATCTCTCTTCCAGAAC	471
QY	163	IleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSer	182
DB	472	ATGGATGCCAAGGAATTTGTATAAATGAACAAGGAGGACTTCTCCGAGCCACCTCCGCC	531
QY	183	TyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThrPhe	202
DB	532	TACACACAGAAAGTGTGTTCGCACCTCAGTTACCTACGAGAA-----AGTTCACTG	585
QY	203	IlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgProAsp	222
DB	586	CTGGCTTATAACACACCTCCCATACAGACCCAGTCTCTCCAGCTGAATGTCAAGGAAGAC	645
QY	223	LeuProTyrGluGlnAlaArgSerAlaTyrThrSerHisSerHisProThrGlnSer	242
DB	646	CCCTCTTATGACTCTCTCAGGAGAGGAGCATGAACAATAATATGAATCTGGCCCTCAAC	705
QY	243	LysAla---ThrGlnProSerSerThrValProLysThrGluAspGlnArgProGln	261
DB	706	AAAGTCTCTCTCTGGAGATCAGACACCATGGCAAGAACACTGAGCAGCGCCCCAG	765
QY	262	LeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGly	281
DB	766	CCAGATCTCTTATCAGATCTCTGGGGCCCAACACGACCGCTAGCAAAACCTTGGGAGTGG	825
QY	282	GlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCys	301
DB	826	CAGATCCAGTGTGGGAGTTTCTCTTGGAACTACTGTCCGACAGCGCCACCGCAGCTGT	885

Alignment Scores:

Pred. No.:	5.86e-113	Length:	1356
Score:	1428.00	Matches:	270
Percent Similarity:	75.93%	Conservative:	50
Best Local Similarity:	63.98%	Mismatches:	92
Query Match:	58.43%	Indels:	10
DB:	29	Gaps:	6

US-09-902-772-2 (1-451) x AY418894 (1-1256)

Qy	35	AlaSerSerSerSerSerGluTyrGlyGlnThrSerLysMetSerProArgValProGlnGln	54
Db	3	GCCTCGGGAGGTCNNGATNCGGGCAGCCCAAGATCAACCCCTCCACCAANNNG	62
Qy	55	AspTrrLeuSerGlnProProAlaArgValThrIleLysMetGluCysAsnProAsnGln	74
Db	63	GAGTGGATCAANNNN---NCNNNAGGCTCAACNNNAAGCGG-----GNNNATGACCAC	113
Qy	75	ValAsnGlySerArgAsnSerProAspAspCysSerValAlaLysGlyLysMetVal	94
Db	114	ATGAATGGATCCAGGGAGCTCCGGTGACCTGCAGCGTTAGCAAAATGACCAAGACTGGTG	173
Qy	95	SerSerSerAspAsnValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---	115
Db	174	GGCGGAGGCGAGTCCACCCCATGAACTACACAGCTATATGGACGAGAAGATGSCCCC	233
Qy	114	ProProProAsnMetThrThrAsnGluArgArgValIleValProAlaAspProThrLeu	133
Db	234	CTCTCTCCCAACATGACCACCAACAGAGAGAGATCATCGTCCCGCAGACCCCAACAGT	293
Qy	134	TrpSerThrAspHisValArgGlnTrpLeuGluTrpAlaValIysGluTyrGlyLeuPro	153
Db	294	TGGACACAGGAGCATGTGAGCAATGGCTGGAGTGGGCCCATAAAGGAGTACACCTTGAT	353
Qy	154	AspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLys	173
Db	354	GAGATCGACACATCCTTTTCCAGAAACATGATGGCAGGAAGTGTGTAAATGAACAAG	413
Qy	174	AspAspPheGlnArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHis	193
Db	414	GAGACTTCTCGTGCCACACCCTCTACACACGGAGAGTGTGTGTCTCACACCTCAGT	473
Qy	194	TyrLeuArgGluArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla	213
Db	474	TACCTCAGGGAA-----AGTTTCACTGGTGCCTTAATACAACTCCCAACACGACCA	527
Qy	214	ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrp	233
Db	528	TCTTCAGACTGAGTGTCAAGAAAGCCCTTCTTATGACTCAGTCAGAAGAGGAGCATGG	587
Qy	234	ThrSerHisSerHisProThrGlnSerLysAlaThrGln---ProSerSerSerThrVal	252
Db	588	GGCAATTAACATGAATCTGGCCTCAACAAAAGTCTCTCCCTGGAGGGGCAACAAACGATC	647
Qy	253	ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer	272
Db	648	AGTAAGAATACAGACCAACGGCCCCAGCCAGATCGGTATCAGATCCTGGGCCCGCACCG	707
Qy	273	SerArgLeuAlaAsnProGlySerGlyGlnIleGlnIleuTrpGlnPheLeuLeuGluLeu	292
Db	708	AGTCGCCTAGNNNNNNNNAAGCGGGCAGATCCAGCTGTGGCAATTTCTCTCGTGGAGCTG	767
Qy	293	LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys	312
Db	768	CTCTCCGACAGCGCAACGCCAGCTGTATCCTCGGAGGGGACCAACGGGGAGTTCAAA	827
Qy	313	MetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMet	332
Db	828	ATGACGGACCCCGATGAGTGGCCAGGCGCTGGGCGAGCGGAAGAACAAGCCCAACATG	887
Qy	333	AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLys	352

compatible sites of a custom modified MCS of the pBlueScript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Alignment Scores:

Pred. No.: 4,55e-100 Length: 726
Score: 1275.00 Matches: 236
Percent Similarity: 99.16% Conservative: 1
Best Local Similarity: 98.74% Mismatches: 0
Query Match: 52.17% Indels: 2
DB: 13 Gaps: 0

US-09-902-772-2 (1-451) x BU334592 (1-726)

Qy 215 GlnArglleThrThrArgProAspLeuProTyGluGlnAlaArgSerAlaTrpThr 234
Db 6 GAGAGAATAACAACAGGCCAGATTTTACCTTATGAGCAAGCAGGAGATCAGCGTGAGC 67
Qy 235 SerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLys 254
Db 68 AGTCAGGCCATCCATCAGTCAGTCAAGAGCTACCAACCATCATCTTCAACAGTGCCCAA 127
Qy 255 ThrGluAspGlnArgProGlnLeuAspProTyGlnlleLeuGlyProThrSerSerArg 274
Db 128 ACAGAGACCAGCGTCCTCAGTTAGATCCTTATCAGATCTTGAGCCGACAGCAGCGGT 187
Qy 275 LeuAlaAsnProGlySerGlyGlnlleGlnLeuTrpGlnPheLeuLeuGluLeuSer 294
Db 188 CTTCCAAATCAGGAGGTGGCAGATACAGTATGGCAGTTCCTACGTGGAGCTTCTGTGC 247
Qy 295 AspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThr 314
Db 248 GACAGCTCCAACTCCAACTGCATCACCTGGAGGGCCAAATGGGAGTTCAAGATGACA 307
Qy 315 AspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyr 334
Db 308 GACCTGATGAAGTGGCTCGCGTGTGGGAGAGGAGAAAGCAACCTAATCATGAACTAT 367
Qy 335 AspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHis 354
Db 368 GACAAACTCAGCGCGCATCTCGCTACTATGACAAATAATATTATGACTAAAGTTTCA 427
Qy 355 GlyLysArgTyrAlaTyrLysPheAspPheHisGlylleAlaGlnAlaLeuGlnProHis 374
Db 428 GGTAAACGCTATGCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCAC 487
Qy 375 ProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHis 394
Db 488 CTTCCAGNATCATCCATGTACAAATACCCATCAGACCTCCCTACATGAGTTCTACCAT 547
Qy 395 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 414
Db 548 GCACACCCCGAGAAGATGAATTTGTAGTCCCATCCCTGCTTTGGCCCGTAACCTCA 607
Qy 415 SerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlylleTyrPr 434
Db 608 TCCAGCTTTTGTGTCCTCCCTTAATCCATACATGGAATTCACCACTGGGAGCATCTACCC 667
Qy 434 cAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 668 CAATACCAAGGTGCCAGTGTCTATATGCCCTTCCATCTTGGCACCTACTAC 720

RESULT 12

CB269538 709 bp mRNA linear EST 20-FEB-2003
LOCUS 1008445 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
DEFINITION cDNA 5', mRNA sequence.
ACCESSION CB269538
VERSION CB269538.1 GI:28444123

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 709)
Yang, R.-Z., Shuldiner, A. and Gong, D.-W.

AUTHORS

EST analysis of human adipose gene expression

TITLE

JOURNAL

COMMENT

Contact: Gong Da-Wei
Division of Endocrinology, Diabetes and Nutrition
University of Maryland
660 Redwood St, HH497, Baltimore, MD 21201, USA
Tel: 410 706 1672
Fax: 410 706 1622
Email: dgong@medicine.umaryland.edu

PCR Primers

FORWARD: CTCGGAGCGCGCATTTGTGTGTG
BACKWARD: AATACGACTCATTAGGGGGAATTGG

Seq primer:

GTGTGATCCGGGAATTC.

Location/Qualifiers

source

1..709

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/sex="Male and Female"

/tissue_type="Adipose"

/clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"

/note="Vector: lambdaTriplex"

ORIGIN

Alignment Scores:
Pred. No.: 5,84e-87 Length: 709
Score: 1123.50 Matches: 218
Percent Similarity: 94.02% Conservative: 2
Best Local Similarity: 93.16% Mismatches: 13
Query Match: 45.97% Indels: 4
DB: 14 Gaps: 1

US-09-902-772-2 (1-451) x CB269538 (1-709)

Qy 204 PheProAsnThrSerValTyrProGluAlaThrGlnArglleThrThrArgProAsnLeu 223
Db 11 TTTCCNAATACATTCTAGTATATCTCTGAGCTACGCAAGAAATACAACTAGGCCAGATT 70
Qy 224 ProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThr---GlnSer 242
Db 71 CCATATGAGCCCCCAGAGAGATCAGCTGGACCGCTCAGGCCACCCACGCCCCAGTCG 130
Qy 243 LysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgProGlnLeu 262
Db 131 AAAGCTGTCTCAACCATCTCTTCCACAGTGGCCCAAACTGAAGACGAGCGTCTCAGTTA 190
Qy 263 AspProTyrGlnlleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGln 282
Db 191 GATCCTTATCAGATTCTTGGACCAACAAGTAGCAGCTTGCAATCCAGGAGTGCCAG 250
Qy 283 IleGlnLeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCysIle 302
Db 251 ATCCAGCTTTGCGAGTTCTCTCTGAGTCTCTGCGACAGCTCCCACTCCAGCTGCATC 310
Qy 303 ThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArg 322
Db 311 ACCTGGGAAGTCAACAAACGGGAGTTCAAGTAGCGGATCCCGACGAGTGCCCGGCGC 370
Qy 323 TrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArg 342
Db 371 TGGGGAGAGCGGAGAGCAACCACTAGTAACTAGCTCAGCGCGCCCTCCGT 430
Qy 343 TyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe 362
Db 431 TACTACTGACAGAACAATCATGACCAAGGTCTTGGAGCGCTACGCCCTACAGATTC 490

```

QY 363 AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMetTyrLys 382
Db 491 GACTTCACGGGATCGCCAGGCCCTCCAGCCACCACCCCGGAGTCATCTCTGACAG 550
QY 383 TyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetAsn 402
Db 551 TACCCCTCAGACCT-CCGTACATGGGCTCCTATACGCGCCACCCACAGAGTGAACCTT 609
QY 403 ValAlaProHisProAlaLeuProValThrSerSerSerPheAlaAlaProAsn 422
Db 610 GTGCGCGCCACCCCTCCAGCCCT-CCGTGACATCTTCCAGTCTTTTGTCTGC-CCAAAC 667
QY 423 ProTyrTrpAsnSerProThrGlyGlyIleTyrProAsnThr 436
Db 668 CCATCTGGATTCACCAACTTGGGGATATACCCCACT 709

RESULT 13
AJ456498 827 bp mRNA linear EST 22-APR-2002
LOCUS AJ456498 riken1 Gallus gallus cdna clone 8m13r2, mRNA sequence.
DEFINITION AJ456498
ACCESSION AJ456498
VERSION AJ456498.1 GI:20266594
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 827)
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
1..827
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="8m13r2"
/cell_type="bursal lymphocyte"
/dev_stages="2-3 weeks old"
/clone_lib="riken1"
/note="TCB inbred strain"

ORIGIN
Alignment Scores:
Pred. No.: 2,01e-86 Length: 827
Score: 1118.50 Matches: 211
Percent Similarity: 78.31% Conservative: 20
Best Local Similarity: 71.53% Mismatches: 43
Query Match: 45.77% Indels: 21
Db: 9 Gaps: 3

US-09-902-772-2 (1-451) x AJ456498 (1-827)
QY 138 HisValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIle 157
Db 5 CACGTGCCCGCAGTGGCTGGATGGCCATAAAGGAGTATGATTAAATGGAGATTGACACC 64
QY 158 LeuLeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGln 177
Db 65 ACCCTCTTCCAAATATGATGGCAAGAGCTCTGCAAAACGAAACAGAGTACTTCTC 124
QY 178 ArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlu 197
Db 125 CGAACCACTCCCTCTACACACAGAGTCTGTGTCTCACTCAGTACTCCTCAGGAA 184
QY 198 ArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIle 217

```

```

Db 185 AGTAGCTCACTGCTGCTCTAC---AATACTCCATCCACACAGAGCTTCCTCAGCTTT 241
QY 218 ThrThrArgProAspLeuProTyrGlnAlaAlaArgSerAlaTyrThrSerHisSer 237
Db 242 GCCACCAAGAGGTCTCTCTGTT----- 265
QY 238 HisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAsp 257
Db 266 -----GCAGGGACACMAAACGTGAACAAGACA-----ACAGAACAG 301
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 302 CAACGGCTCAGCCAGATCCCTATCAAAATCTGGGGCCCACTAGTAGTCGTCCTGCCAAT 361
QY 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuSerAspSerSer 297
Db 362 CTGGGAGTGGGAGATACAACTATGGCAGTTCCTCTCGAGTTGCTGCGACAGTTCC 421
QY 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 422 AACGCCAGCTGTATCACATGGGAGGACCAATGGGGAATTCAGATCACAGACCAGAT 481
QY 318 GluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 482 GAAGTGGCAGCGCGCTGGGAGAACGCAAAAGCAAGCCCAATGATATGACAAAGCTG 541
QY 338 SerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 542 AGCGAGCGCTTCGATACTACTATGATGAAGACATTATGACCAAGTGCATGGCAAAAG 601
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 602 TATGCTACAAATTGACTTTTCATGGCATTCGCCAGGCTCTCCAGCCTCATCCCACTGAA 661
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 662 TCATCAATGTACAGTATCCATCAGATCTCTCTCATGCGCTTCTTACCATGCCACCAG 721
QY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhe 417
Db 722 CAGAAGGTGAACTTGTATGCCCAACACCTCTTCTTATGCGCTGTACATCATCCAGTTTC 781
QY 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIle 432
Db 782 TTCGAGCAGCGCTCACCTTATTGGACCTCCCTGCTGGGAAGCATT 826

RESULT 14
AK083900 2420 bp mRNA linear HTC 20-SEP-2003
LOCUS AK083900
DEFINITION Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length
enriched library, clone:D130057B15 product:Friend leukemia
integration 1, full insert sequence.
ACCESSION AK083900
VERSION AK083900.1 GI:26350832
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

```

```

3  Saibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
   Kono,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M.,
   Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
   Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
   Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
   Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
   Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
4  The RIKEN Genome Exploration Research Group Phase II Team and the
   FANTOM Consortium.
   Functional annotation of a full-length mouse cDNA collection
   Nature 409, 685-690 (2001)
5  The FANTOM Consortium and the RIKEN Genome Exploration Research
   Group Phase I & II Team.
   Analysis of the mouse transcriptome based on functional annotation
   of 60,770 full-length cDNAs
   Nature 420, 563-573 (2002)
6  (bases 1 to 2420)
   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
   Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
   Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
   Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
   Katon,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
   Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
   Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
   Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
   Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
   Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,P., Takaku-Akahira,S.,
   Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
   Muramatsu,M. and Hayashizaki,Y.
   Direct Submission
   Submitted (15-APR-2002) Yoshihide Hayashizaki, The Institute of
   Physical and Chemical Research (RIKEN), Laboratory for Genome
   Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
   RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
   Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
   URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
   Fax:81-45-503-9216)
7  cDNA library was prepared and sequenced in Mouse Genome
   Encyclopedia Project of Genome Exploration Research Group in Riken
   Genomic Sciences Center and Genome Science Laboratory in RIKEN.
   Division of Experimental Animal Research in Riken contributed to
   prepare mouse tissues.
   Please visit our web site for further details.
   URL:http://genome.gsc.riken.go.jp/
   URL:http://fantom.gsc.riken.go.jp/.
   Location/Qualifiers
     1. 2420
        /organism="Mus musculus"
        /mol_type="mRNA"
        /db_xref="FANTOM DB:D130057B15"
        /db_xref="MG1:2419853"
        /db_xref="taxon:10090"
        /clone="D130057B15"
        /tissue types="spinal ganglion"
        /clone_lib="RIKEN full-length enriched mouse cDNA library"
        /dev_stages="12 days embryo"
        cl. 854
        /note="unnamed protein product; Friend leukemia
        integration 1 (MGD|MG1:95554, GB|NM_008026, evidence:
        BLASTN, 99%, match=1671)
        Putative"
        /codon_start=3
        /protein_id="BAC39053.1"
        /db_xref="GI:26350833"
8  CDS
   /translation="MNKEPLRATSAVNTVELSHLSYRESSLLAYNTTSHTDQSSR
   LNKEDSYDVRGANNMNSGLKSPFLGSGQTGKNTQRPQDPYQILGPTSS
   RLNPSSGQQLWQFQLELLSDSANASCIITWETNGFMDDEVARWGERKSPN
   MNTDKLSRLRYDDKNIMTKVHGKYAKDFHGAQAOLPHPTETSMYKYPSDISY
   MPSTYHARQQKVNFPVPSHPMPVTSSFFGAASQYWTSTAGTIPNPSVPRHPTHP
   SHLSYY"
   ORIGIN
   Alignment Scores:
   Pred. No.: 3,32e-80 Length: 2420
   Score: 1055.00 Matches: 196
   Percent Similarity: 80.00% Conservative: 32
   Best Local Similarity: 68.77% Mismatches: 51
   Query Match: 43.17% Indels: 6
   DB: 11 Gaps: 3
   US-09-902-772-2 (1-451) x AK083900 (1-2420)
   QY 171 MetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSer 190
   DB 3 ATGAACAAGGAGGAGCTTCCTCCGAGCCACCTCCGCTTACACACAGAGAAGTCTGTGTGCG 62
   QY 191 HisLeuHisTyrLeuArgGluArgGlyAlaThrPheIlePheProAsnThrSerValTyr 210
   DB 63 CACCTCAGTTACCTCAGGGAA-----AGTTTCATCTGCTGGCTATACACACCTCCCAT 116
   QY 211 ProGluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArg 230
   DB 117 ACAGACCAAGTCTCCTCAGCACTGAATGTCAAGGAGAGACCTCTTATGACTCTGTGAGGAGA 176
   QY 231 SerAlaThrThrSerHisSerHisProThrGlnSerLysAla---ThrGlnProSerSer 249
   DB 177 GGAGCATGGAAACAATAATATGAAGCTGGCTGGCTCAACAAAGAGTCTCTCTTGGAGGATCA 236
   QY 250 SerThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGly 269
   DB 237 CAGACCATGGGCAAGAACACACTGAGCAGCGGCCCCCAGCAGATCCTTATCAGATCTGCGGG 296
   QY 270 ProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrGlnPheLeu 289
   DB 297 CCAACCAAGCCGCTAGCAAAACCTGGGAGTGGGAGATCCAGCTGTGGCAGTTTCTC 356
   QY 290 LeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGly 309
   DB 357 CTGGACTACTGTCCGACAGCGCCACCGCAGCTGTATCACTGGAGGGGACCAACGCG 416
   QY 310 GluPheLysMetThrAspProAspGluValAlaArgTyrGlyGluArgLysSerLys 329
   DB 417 GAGTTCAAAATGACGACCCCTGATGAGTGGCCAGCGGCTCGGAGAGCGGAGGAGCAAG 476
   QY 330 ProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIle 349
   DB 477 CCCAACATGAATTATGACACAGCTGAGCGCGGCTCCGATCTACTACTATGACAAAACATT 536
   QY 350 MetThrLysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGln 369
   DB 537 ATGACCAAAGTGCATGGCAAAAGGTATGCTTACAAAGTTTGACTTCCATGGCATTTGCCAG 596
   QY 370 AlaLeuGlnProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyr 389
   DB 597 GGCCTTGCAGGCACATCCCAACAGAGACATCCATGATACAGTATCCCTCTGATATCTCTAC 656
   QY 390 MetSerSerTyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAla 409
   DB 657 ATGCGCTTCTACCATGCCATCAACAGAAAGGTGAACCTTGTCCCGCTCTCACCCATCTCC 716
   QY 410 LeuProValThrSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThr 429
   DB 717 ATGCGCTGTCACTCTCTCAGCTTCTTTGGAGCAGCATCAATATCTGAGACCTCCCTCCACT 776
   QY 430 GlyGlyIleTyrProAsn-----ThrArgLeuProAlaAlaHisMetProSerHis 446
   DB 777 GCTGGGATCTATCCAAACCCAGTGTGCTCCCGCCCATCTTAACACCCAGCTGCTTACACAC 836

```

```

QY      447 LeuGlyThrTyrTyr 451
Db      837 TTAGGCGAGCTACTAC 851

RESULT 15
BM456833
LOCUS
DEFINITION 6404041 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583753
5', mRNA sequence.
ACCESSION BM456833
VERSION   BM456833.1 GI:18505873
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1045)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-k@mail.nih.gov
           Tissue Procurement: ATCC
           CDNA Library Preparation: Life Technologies, Inc.
           CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LAM2347 row: f column: 10
           High quality sequence stop: 569.

FEATURES
     source
       1..1045
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="IMAGE:5583753"
         /tissue_type="embryonal carcinoma, cell line"
         /lab_host="DH10B (phage-resistant)"
         /clone_lib="NIH_MGC_92"
         /note="Organ: testis; Vector: pCMV-Sport6; Site 1: NotI;
           Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
           Average insert size 2.5 kb. Library enriched for
           full-length clones and constructed by Life Technologies.
           Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:      1,82e-79      Length:      1045
Score:          1040.00      Matches:    214
Percent Similarity: 75.24%      Conservative: 17
Best Local Similarity: 69.71%      Mismatches: 33
Query Match:    42.55%      Indels:    44
DB:              12          Gaps:       8

US-09-902-772-2 (1-451) x BM456833 (1-1045)

QY      7 GluAlaLeuSerValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26
Db      190 GAAGCCTTATCAGTTGTGAGTGAGGACAGTCGTTGTGTAGTGTGCTACGGAACGCCA 249

QY      27 HisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyrGlyGlnThrSerLys 46
Db      250 CACCTGGCTAAGACAGAGATGACCGCTCTCTCCAGCGACTATGCACAGACTTCCAAAG 309

QY      47 MetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArgValThrIle 66
Db      310 ATGAGCCACACCGCTCCCTCAGCAGGATGGCTGTCTCAACCCCGACGACGGGTCAACATC 369

QY      67 LysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspCysSer 86
Db      370 AAAATGGATGTAACTTAGCCAGGTGAATGGCTCAGAGAACTCTCTCTGATGATGCAGT 429

```

```

QY      87 ValAlaLysGlyGlyLysMetValSerSerAspAsnValGlyMetAsnTyrGlySer 106
Db      430 GTGGCCAAAGCGGGAGATGTTGGGAGCCACACACCGTTGGATGAATCACTCGGCAGC 489

QY      107 TyrMetGluGluLysHisIleProProAsnValThrThrAsnGluArgValIle 126
Db      490 TACATGGAGGAGAGACACATGCCACCCCAACATGATGACCAAGACGAGCGGAGATTATC 549

QY      127 ValProAlaAspProThrLeuTyrSerThrAspHisValArgGlnTyrLeuGluTyrAla 146
Db      550 GTGCCAGCAGATCTTACGCTATGGAGTACAGACCATGTGGGCGAGTGGCTGGAGTGGCGG 609

QY      147 ValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys 166
Db      610 GTGAAGAATAATGGCTTCCAGACGTCACATCTTGTATTATCCAGAACATCGATGGGAG 669

QY      167 GluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAsp 186
Db      670 GACTGTGCAAGATGACCAAGAGCAGCTTCAGAGGCTCACCCCGAGTACACCGCCGAC 729

QY      187 IleLeuLeuSerHisLeuHisTyrIleuArgGluArgGlyAlaThrPheIlePheProAsn 206
Db      730 ATCCTTCTCTCACATCTTCACTACCTCAGAGAGACTCTCTTCCACATTTGACTCCAGAT 789

QY      207 ThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro----- 221
Db      790 -----GATGTTGATAAAGC-CTTACAAACTCTCCCGGTTAATGCATGCT 833

QY      222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTyrThrSerHisSerHis 238
Db      834 AGAAACACCGATTATACCATATGGACCCCGGAGATCAGCTGGACCGGTCAACGTCAC 893

QY      239 ProThrGlnSerLysAlaThrGlnProSerSerSerThrVal-----Pro 253
Db      894 CCC-----CAAGCCCAAGTCGAAAGCTGTTCAAACATTTCTCCCTCAAGGGCCC 944

QY      254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273
Db      945 AAAACTGAA-----AGACCCAGC-----GGCCCTCAGGT--- 974

QY      274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuLeuLeu 293
Db      975 -----TAAATCCC-----TTAATCACAATCCTT 998

QY      294 SerAspSerSerAsnSerAsn 300
Db      999 TGGGAACCCCCACATATAAAC 1019

```

Search completed: May 27, 2004, 23:56:42
Job time : 3667 secs